JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXXVI NOVEMBER, 1942 Number 3

RELATIONSHIP OF ELEMENTARY-SCHOOL ACHIEVEMENT TESTS TO ACHIEVEMENT TESTS TAKEN IN THE SECONDARY SCHOOL

ARTHUR E. TRAXLER and MARGARET S. SELOVER

Editor's note: There have been many attempts to find tests of prognostic value. The authors find that elementary achievement tests have prognostic value in predicting secondary school achievement particularly in the language subjects.

THE most important single argument in favor of the use of achievement tests is the hypothesis that they enable one to predict the achievement of a pupil at some later time better than the prediction can be made on the basis of chance, or hunches, or a casual subjective appraisal of the pupil. If this hypothesis is correct, tests can help a school undertake a reasonable, objective program of instruction and guidance for each pupil from the time of entrance.

The usefulness of tests for prediction over a period of a few weeks or even a few months is fairly well established as a result of both objective studies and the experience of many different teachers and schools. Their predictive value for periods longer than one year is less well established, because of the difficulty of following up pupils from year to year. This is true especially where secondary-school tests are used to predict achievement-test results in college or where elementary-school tests are employed to predict achievement on tests taken in the secondary school.

The fact that many independent schools administer achievement tests annually and have the tests scored at the Educational Records Bureau, where a record of the results is kept on file, has created a reservoir of test data which may be used in studying relationships over a period of several years. This article reports such a study in which a comparatively small amount of data was used. It is indicative of a type of study that can be made on a much more extensive basis if and when sufficient staff time and funds for research

become available. It deals with the correlations between achievement tests taken in the elementary school and certain achievement tests used in the secondary school.

SOURCE OF THE DATA

Most of the elementary-school test scores used in this study were obtained with the New Stanford Achievement Test. In some instances, the Metropolitan Achievement Test was given instead of the New Stanford, but these results were expressed in terms of New Stanford scores by means of the publisher's table of equivalent scores on the two tests. The elementary-school test data were obtained in Grades V to VIII, inclusive.

The secondary-school test results were obtained with the Cooperative tests in English, literary acquaintance, literary comprehension, elementary algebra, intermediate algebra, and place geometry. These tests were distributed throughout Grades IX to XII.

One of the conditions of the study was that the pupils involved must have transferred from an elementary school to an entirely different secondary school, for it is in such circumstances that test results are especially needed for predictive purposes. Moreover, it was felt that this condition would provide a more severe basis of appraisal of the tests than would be true where the pupils remained within the same school system and were perhaps given a more or less constant type of instruction.

The fact that transfer from one school to another was made one of the conditions of the study and the further fact that it was necessary to follow up the pupils over a long period limited the number of available cases that were finally obtained from the sampling of records examined. A total of thirty-six long-time individual cumulative records formed the basis of the study, but the number of cases included in some of the correlations was considerably less than that, for some of the records did not extend over the entire range of grades from V to VII and not all the tests were taken by every pupil.

QUESTIONS RAISED

This article is concerned with the following questions:

1. Is there a significant degree of relationship between the achievement tests taken in the elementary school and those taken in the secondary school?

2. Is the size of the correlations influenced by the time interval between the dates of administration?

3. Is there a difference between the predictive value of the linguistic and the mathematical tests?

THE RESULTS

The correlations of scores on achievement tests taken in Grades V, VI, VII, and VIII with scores on Cooperative achievement tests taken in Grade IX are shown in Table I. The correlations fall within the range .32 to .78. All the correlations between average reading scores on the Cooperative English Test and between average reading scores and scores on the vocabulary part of the English test are statistically significant and most of them are rather high. All those for spelling in the elementary school and spelling in the secondary school are also significant and fairly high. Three of the four correlations between the language usage test in the elementary school and the usage part of the Cooperative English Test are significant. The one that is not is for the shortest time interval, one year.

TABLE I

CORRELATIONS BETWEEN ACHIEVEMENT TEST* SCORES IN GRADES V–VIII AND SCORES ON COOPERATIVE ACHIEVEMENT TESTS TAKEN IN GRADE IX

Tests	Grade when achievement tests* were taken											
	V			VI			VII			VIII		
	N	r	PE	N	r	PE	N	r	PE	N	r	PE
Av. Reading vs. Coop. Eng. (Tot.)	21	.50	±.11	30	.78	±.05	36	.72	±.05	15	.69	±.09
Av. Reading vs. Coop. Eng. Vocab.	21	.71	±.07	30	.69	±.07	35	.75	±.05	15	.74	±.08
Language Usage vs. Coop. Eng. Usage	21	.53	±.11	30	.53	±.09	36	.57	±.08	15	.36	±.15
Spelling vs. Coop. Eng. Spelling	21	.66	±.08	30	.68	±.07	34	.71	±.06	28	.69	±.07
Literature vs. Coop. Lit. Acq.	17	.56	±.11	19	.42	±.13	23	.53	±.10	19	.47	±.12
Av. Arithmetic vs. Coop. Elem. Algebra	18	.42	±.13	27	.50	±.10	32	.43	±.10	28	.32	±.12

* Achievement tests were Metropolitan or New Stanford, and scores were expressed in terms of New Stanford scores.

Of the four correlations between the elementary-school literature test and the Cooperative Literary Acquaintance Test, two are clearly significant and two are probably so.

The correlations between average arithmetic in the elementary school and the Cooperative Elementary Algebra Test in Grade IX are the only set in which none of the coefficients is above .50. However, two of the correlations are significant statistically and one is probably so.

Clearly, there is some positive relationship between the achievement tests administered in Grades V, VI, VII, and VIII and the achievement tests given in Grade IX. Most of the correlations are only fairly high and some are rather low, but very high correlations are not to be expected, since the elementary-school and secondary-school tests do not measure exactly the same abilities. The correlations would no doubt be higher if appropriate batteries of the same series of achievement tests were used from Grade V upward through the secondary school.

On the whole, the verbal achievement tests used in Grades V to VIII, inclusive, predict achievement in English and literary acquaintance at the ninth-grade level better than the test of arithmetic achievement predicts success on the elementary algebra test.

As far as the data in Table I are concerned, the tests given in Grade VII tend to predict achievement scores in Grade IX better than the tests in the other elementary-school grades. In general, there is no significant difference in the accuracy with which the tests administered in Grades V, VI, and VIII predict ninth-grade achievement.

Further evidence about the relationship of elementary-school and secondary-school achievement tests is provided by Table II, which shows correlations between scores on achievement tests given in Grade VII and achievement tests administered in Grades IX–XII.

Most of the correlations for the verbal tests in Table II vary from .5 to .8, although several are considerably below this range and one is above it. All the correlations between average reading in Grade VII and Cooperative English total score, between average reading and Cooperative English vocabulary, between average reading and Cooperative literary comprehension, and between spelling in Grade VII and the spelling part of the Cooperative English Test have statistical significance. The correlations of the language usage scores in Grade VII with the scores on the usage part of the Coopera-

TABLE II

CORRELATIONS BETWEEN ACHIEVEMENT TEST[x] SCORES IN GRADE VII AND SCORES ON COOPERATIVE ACHIEVEMENT TESTS TAKEN IN GRADES IX–XII

Tests	Grade when Cooperative tests were taken											
	IX			X			XI			XII		
	N	r	PE	N	r	PE	N	r	PE	N	r	PE
Av. Reading vs. Coop. Eng. (Tot.)	36	.72	±.05	36	.69	±.06	33	.67	±.07	34	.63	±.07
Av. Reading vs. Coop. Eng. Vocab.	35	.75	±.05	35	.84	±.03	33	.62	±.07	33	.69	±.06
Language Usage vs. Coop. Eng. Usage	36	.57	±.08	36	.35	±.10	33	.21	±.11	34	.19	±.11
Spelling vs. Coop. Eng. Spelling	34	.71	±.06	34	.57	±.08	31	.52	±.09	32	.65	±.07
Literature vs. Coop. Lit. Acq.	23	.53	±.10	23	.35	±.12	18	.47	±.12	17	.50	±.12
Av. Reading vs. Coop. Lit. Compr.							15	.51	±.13	17	.71	±.09
Av. Arithmetic vs. Coop. Elem. Algebra	32	.43	±.10									
Av. Arithmetic vs. Coop. Inter. Algebra				18[y]	.39	±.14						
Av. Arithmetic vs. Coop. Plane Geometry							15[z]	.24	±.16			

x Achievement tests were Metropolitan or New Stanford, and scores were expressed in terms of New Stanford scores.

y Cooperative Intermediate Algebra taken in Grades IX-XI.

z Cooperative Plane Geometry taken in Grades X-XII.

tive English Test are low, except in the case of usage scores obtained in Grade IX. The correlations between the literature test taken in Grade VII and the literary acquaintance test in Grades IX to XII are not very high, but with the exception of the one for the tenth-grade level, they tend to be significant.

All three correlations between average arithmetic scores at the seventh-grade level and scores on the Cooperative mathematics are below .5. The

correlation between arithmetic and elementary algebra is the only one of the three that is more than four times its probable error.

The general indications of the correlations in Table II are similar to those of the data in Table I. There is, on the average, considerable positive relationship between the results of the elementary-school and the secondary-school achievement tests. The correlation tends to be closer for the verbal tests than for the mathematical ones. There is some tendency for the correlations to decrease with increase in time interval, but this tendency is not marked.

In addition to information about the predictive value of tests in particular subjects, data on the usefulness of total achievement scores in the elementary school for the prediction of composite achievement in the secondary school are of considerable interest. Some information on this point is given in Table III, which shows correlations between total achievement scores in Grades V to VIII and combined scores on the Cooperative tests in English and elementary algebra administered in Grade IX.

TABLE III

CORRELATIONS BETWEEN TOTAL AVERAGE SCORES ON THE NEW STANFORD ACHIEVEMENT TEST[x] TAKEN IN GRADES V THROUGH VIII AND SCALED SCORES ON COOPERATIVE ENGLISH AND COOPERATIVE ELEMENTARY ALGEBRA, COMBINED, TAKEN IN GRADE IX

Grade of New Stanford Test	English plus Algebra		
	N	r	PE
V	18	.58 ± .11	
VI	27	.63 ± .08	
VII	30	.73 ± .06	
VIII	11	.71 ± .10	

x Some tests were Metropolitan rather than New Stanford Achievement Tests, but all scores were expressed in terms of New Stanford scores.

The correlations vary from .56 for the achievement tests given in Grade V to .73 for the tests administered in Grade VII. These correlations are fairly high, although they are not high enough to provide a very reliable basis for individual prediction. They would probably be higher if the tests taken by the pupils in the secondary school included all the kinds of achievement covered by the elementary-school tests.

CONCLUSIONS

The number of cases in the study is not large enough to warrant definite statements about the relationships between elementary-school and secondary-school achievement tests, but a careful inspection of the forty-seven correlations reported suggests the following conclusions:

1. Achievement tests given in the elementary school have some value for the prediction of achievement in the secondary school, one, two, three, or four years later.

2. In general, there is some evidence that the predictive value of elementary-school achievement tests tends to decrease with increase in the time interval between the elementary-school and the secondary-school tests, but for the intervals one to four years, the decrease seems to be slight.

3. The prognostic value of elementary-school achievement tests in the linguistic field appears to be somewhat greater than that of tests in the field of mathematics.

A SIMPLE METHOD OF SCORING AND INTERPRETING SEQUENTIAL RESPONSES

A. C. ROSANDER
War Production Board
Washington, D. C.

Editor's note: In many instances test constructors have desired to employ sequential responses but have found this sort of response difficult to grade. The author presents a simple method of scoring such tests.

A NOT uncommon problem in psychological and educational measurement is that of scoring and interpreting responses to questions or situations which require that the answers be given in a specific order or sequence. A whole series of such questions are employed in the social sciences, especially in history. They require the arranging in chronological order of a series of important movements, significant events, famous men, political documents, institutional origins. The same general problem is encountered in the intelligence test item in which actual weights are to be arranged in order of magnitude, and in trade tests wherein the subject is asked to arrange in order several specimens of wood or metal according to degree of hardness. Another example is found in the problem of arranging specimens of composition, of drawings, of literature, and of handwriting in order of merit. Still another example is found in an automobile driver's test in which the nearness of several lights is to be estimated.

In situations of this type wherein rank order is a measure of discriminal ability, the concept of inversion of rank order may be used to indicate the extent of this ability. Assume that on the basis of some independent criterion such as weight, distance, time sequence, or expert judgment, the true or accepted order of the several stimuli is

1, 2, 3, ----------, n-1, n

Perfect discrimination will mean that the individual reproduces exactly this order of position. The greatest departure from this order will be the inverse arrangement

n, n-1, ----------, 3, 2, 1

This is perfect discrimination also but because of misunderstanding of directions or terms, or confusion in perception, the individual has arranged them in the reverse sequence. Midway between these two sequences is the

group which constitutes a random arrangement of the stimuli such as one would obtain by making drawings from Tippett's or Fisher's random numbers. It should be pointed out that the concept of inversions rules out the use of tied or fractional ranks.

Inversions of order. Assume that the problem is to score a history question in which five events are to be arranged in chronological order; assume the events are designated b, d, a, c, and e, and that their time sequence from the earliest to the last is 1, 2, 3, 4, 5. If a person interchanges any pair of events such as indicated by the underlining in each of the following examples

1 3 2 4 5 2 1 3 4 5 1 2 4 3 5 1 2 3 5 4

this constitutes one inversion of order, and is indicative of the first degree of confusion possible in this test question. It is assumed that such confusion of a pair of positions means the same psychologically at whatever position in the sequence it takes place.

If the individual confuses the position of one or two different elements then he may return such ranks as the following, each representing two inversions and a higher degree of confusion:

1 3 4 2 5 3 1 2 4 5 1 2 4 5 3 1 3 2 5 4

In this case the displacement of one element two positions is taken to be equivalent to the displacement of two elements one position.

In this particular example the maximum number of inversions would come from the one arrangement 5, 4, 3, 2, 1, which represents ten inversions of rank. Examples of sequences of five elements having inversions between 3 and 10 in number are as follows:

3 inversions	2, 1, 5, 3, 4	7 inversions	3, 5, 4, 1, 2
4 inversions	1, 3, 5, 4, 2	8 inversions	4, 3, 5, 2, 1
5 inversions	2, 4, 5, 1, 3	9 inversions	5, 4, 3, 1, 2
6 inversions	3, 5, 1, 4, 2	10 inversions	5, 4, 3, 2, 1

The mean number expected on the basis of chance is 5; this is the measure of chance arrangement of the five stimuli. Taking this as the zero point in discrimination we may establish for five stimuli the following scale of discrimination:

Number of Inversions	Discriminal scale
0	5
1	4
2	3
3	2
4	1
5	0
6	—1
7	—2
8	—3
9	—4
10	—5

Properties of inversion distributions. It can be shown that for any number of ranks n there is a symmetrical distribution whose mean and variance are functions only of n, and that as n increases, the distribution approaches rapidly the normal probability curve.[a] The frequency distributions of all possible inversions for ranks from 2 through 6 are as follows:

Number of Inversions	Frequency of inversions for n elements				
x	n=2	n=3	n=4	n=5	n=6
0	1	1	1	1	1
1	1	2	3	4	5
2		2	5	9	14
3		1	6	15	29
4			5	20	49
5			3	22	71
6			1	20	90
7				15	101
8				9	101
9				4	90
10				1	71
11					49
12					29
13					14
14					5
15					1
Total (n!)	2	6	24	120	720

Each distribution can be constructed from the one preceding it by merely adding horizontally the $n-1$ frequency distributions repeated n times but staggered one position of the variate x; for example the distribution for $n=4$ can be derived from that for $n-3$ in the following manner:

[a] These distribution characteristics were first developed and used by the writer in 1937 as the basis of tests of significance for ranked data.

Number of inversions x	From n = 3				Horizontal sum gives n = 4
0	1				1
1	2	1			3
2	2	2	1		5
3	1	2	2	1	6
4		1	2	2	5
5			1	2	3
6				1	1
Sum	6	6	6	6	24

The minimum number of inversions in any set of n orders is zero, while the maximum number is

$$x_{max} = \frac{n(n-1)}{2}$$

The arithmetic mean of the distribution of inversions for n ranks is

$$M_x = \frac{n(n-1)}{4}$$

Since the distribution is symmetrical the arithmetic mean is also equal to the median and the mode. The variance of the distribution is

$$\sigma^2_{\bar{x}} = \frac{n(n-1)(2n+5)}{72}$$

Hence the sampling variance of a sample mean based on N cases is

$$\sigma^2_{\bar{x}} = \frac{\sigma_x^2}{N} = \frac{n(n-1)(2n+5)}{72N}$$

The first three are exact expressions of the characteristics of the inversion population. In computing the sampling variance of the mean we know what σ^2 is, and hence we do not need to estimate it from the sample. N is the number of sets of ranks for each of which the number of inversions has been counted. These distributions are symmetrical about the mean M, and approach the normal curve rapidly so that when n is 6 the correspondence to the normal curve is quite close. The basic constants for these distributions for values of n from 2 to 10 inclusive are:

n	x_{max}	M_x	$12\sigma_x^2$
2	1	0.5	3
3	3	1.5	11
4	6	3	26
5	10	5	50
6	15	7.5	85
7	21	10.5	133
8	28	14	196
9	36	18	276
10	45	22.5	375

Method of scoring. The method of scoring is simply that of using the number of inversions as an index of discriminal ability. Since a low number of inversions means high discrimination based on the criterion, it is more in line with usual practice to reverse the scales so that the higher scores mean greater degrees of discrimination. Two types of scoring might be used. One would be the discriminal score: the number of inversions minus the population (chance) mean. The merit of these scores is the psychological sense which can be attached to them. They lead, however, to negative values which, although they do have real meaning, may be objected to. To avoid negative scores we suggest the use of a score based upon the maximum number of inversions (x_{max}) minus the obtained number of inversions (x). The three scoring methods suggested are:

1. Number of inversions: x
2. Discriminal score: $M - x$
3. Non-negative scores: $x_{max} - x$

The last two methods have a perfect negative correlation with the number of inversions, and a perfect positive correlation between themselves. To illustrate, assume the following case of 5 stimuli each ranked by ten persons; the correct order is assumed to be 1, 2, 3, 4, 5.

Person	Rank Order	Inversions x	Discriminal score (5-x)	Score (10-x)
A	1 3 2 4 5	1	4	9
B	1 2 3 4 5	0	5	10
C	2 3 5 1 4	4	1	6
D	1 2 3 5 4	1	4	9
E	1 4 3 2 5	3	2	7
F	3 1 2 4 5	2	3	8
G	2 5 3 1 4	5	0	5
H	1 3 2 5 4	2	3	8
I	2 1 4 3 5	2	3	8
J	2 3 1 5 4	3	2	7
Sum		23	27	77
Mean ($\bar{x}$)		2.3	2.7	7.7
Expected mean based on chance		5	0	5

The frequency distributions indicate that for questions or arrangements of this type, four or more stimuli (elements) should be employed in order to reduce the probability of obtaining high scores by chance. When n is 4, the probability of getting a perfect score by chance is 1 in 24; when n is 5 this is reduced to 1 in 120; while if n is 6 it is reduced to 1 in 720. Using 5 or 6 elements therefore will make it exceedingly difficult to guess a high

score. The use of five elements has some merit not only because of the small probability of guessing a high score, but because of the convenient range and size of the scores.

Three hypotheses to test. In any problem of this kind the first hypothesis is whether the discriminal ability of the group is departing significantly from chance; to do this we test the null hypothesis that the sample mean equals the population mean, or that $\bar{x} = M$. In the example cited the sample mean of the group $\bar{x}$ is 2.3 inversions whereas the expected number based on chance, the population mean, is 5. Is the difference of 2.7 the discriminal score of the group, to be expected as a usual deviation? When n is 5 and N is 10 the standard error of the sample mean is 0.65; hence the value of 2.3 is $\frac{2.7}{.65}$ or about 4 standard deviations below the true mean. The probability of obtaining such a deviation by chance is so exceedingly small that we reject the null hypothesis that a sample mean of 2.3 inversions could have readily arisen from a population with a mean of 5 and a variance of 50/12.

A second hypothesis assumes that the variance of the sample is equal to the variance of the population. In our example the problem is to test whether the variance of the 10 observations could have arisen by chance from a population with a known variance of 50/12. Since we know the population variance σ^2, we can use the chi-square test of variance and refer the calculated value to a chi-square table, using $N - 1$ degrees of freedom.

$$X^2 = \frac{N s^2}{\sigma^2}$$

and

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{N - 1}$$

This expression of chi-square is applicable to the number of inversions because we know the true population variance σ^2; it is not applicable to cases where σ^2 is unknown. The *estimated* variance, s^2, from the sample is 2.23, N is 10, while the known true population variance, σ^2, is 50/12. This gives a value of chi-square of 5.35 which for 9 degrees of freedom has a probability of being exceeded about 80 percent of the time. Hence the hypothesis of homogeneity of the sample variance relative to the population variance is not rejected.

A third hypothesis to test is the goodness of fit of the *obtained* inversion frequency distribution relative to the *expected* inversion distribution based on theory. The appropriate test here is the usual chi-square test for goodness of fit. It will not be applied to the present problem because at least 100 values of N, and preferably more, should be available in order to define the distribution.

Applications. In cases where the ability to discriminate the order of a set of elements is influenced by instruction, such as would be illustrated in history on the one hand and literary appreciation on the other, the deviation of the scores of individuals and of groups from chance may be used as a measure of the effectiveness of the instruction. By giving tests of this type before and after a course, the instructor can discover to what extent, in terms of the criterion, the members of the class change their discriminal scores.

Since discriminal scores are affected by stimulus differences, the method described herein may be used to study the effects of varying stimulus differences, and in general to investigate the problem of the difference limen and that of just noticeable differences. Since it is possible by means of the inversion distributions to measure departure from zero discrimination it offers a statistical definition of what constitutes a just noticeable difference.

The use of the inversion distributions in cases where there are 4 elements and 5 elements to arrange in some specified order will be discussed in detail. In constructing test questions based on either 4 or 5 elements it is necessary (1) to know the correct order, (2) to mix the order in which the elements are presented in each of the several questions. Finally each of the 24 possible orders of 4 elements and the 120 possible orders of 5 elements must be scored. In order to facilitate this entire process we present nonnegative scores for every possible order of 4 elements and of 5 elements, given a specified correct order.

In Table I the 24 possible permutations of elements a, b, c, and d are shown in the stub, while the headings represent 10 possible correct orders. The entries in the table are the scores of the various response orders for each of the correct orders.

This table would be used in the following manner. Each element under a question would be lettered in the usual manner: a, b, c, and d. The correct orders would be varied from question to question; 10 possibilities are given. The elements in the first question would be lettered a, b, c, and d

TABLE I

SCORING KEY FOR FOUR ELEMENTS, ABCD

Scores for each of the 24 possible responses based on each of ten correct orders

Response order	Correct orders									
	cadb	adbc	dacb	dbca	cbda	bdac	bcad	cdab	acbd	badc
abcd	3	4	2	1	2	3	4	2	5	4
abdc	2	5	3	2	1	4	3	1	4	5
acbd	4	3	3	0	3	2	3	3	6	3
acdb	5	4	4	1	2	1	2	4	5	2
adbc	3	6	4	3	0	3	2	2	3	4
adcb	4	5	5	2	1	2	1	3	4	3
bacd	2	3	1	2	3	4	5	1	4	5
badc	1	4	2	3	2	5	4	0	3	6
bcad	3	2	0	3	4	3	6	2	3	4
bcda	2	1	1	4	5	4	5	3	2	3
bdac	0	3	3	4	3	6	3	1	2	5
bdca	1	2	2	5	4	5	4	2	1	4
cabd	5	2	2	1	4	1	4	4	5	2
cadb	6	3	3	2	3	0	3	5	4	1
cbad	4	1	1	2	5	2	5	3	4	3
cbda	3	0	2	3	6	3	4	4	3	2
cdab	5	2	4	3	4	1	2	6	3	0
cdba	4	1	3	4	5	2	3	5	2	1
dabc	2	5	5	4	1	4	1	3	2	3
dacb	3	4	6	3	2	3	0	4	3	2
dbac	1	4	4	5	2	5	2	2	1	4
dbca	2	3	3	6	3	4	3	3	0	3
dcab	4	3	5	4	3	2	1	5	2	1
dcba	3	2	4	5	4	3	2	4	1	2

but the correct order might be cadb (see the first column in Table I). The score for each of the 24 possible orders in response to this arrangement is given in the first column. For example, a response of abcd to this specific question would be scored 3, a response of adcb would be scored 4. The order cadb, the correct one, is scored 6; while the reverse order, bdac, is scored 0. The mean score expected by chance on the basis of 4 elements is 3; scores above 3 indicate positive discrimination while those below 3 indicate inverse or negative discrimination.

The scores to use for nine other correct orders are given; nine arrangements are probably enough for usual practice. If it is not, scores can be computed for other correct orders. In every case the score is the maximum number of inversions, 6 in this case, minus the number of inversions found by comparing any given response with the correct one.

Table II gives the scores to use for each of the 120 possible orders of 5 elements abcde, based on four assumed correct arrangements; dbaec, beadc, bceda, and ebacd. These four orders can be increased to eight correct orders

TABLE II

SCORING KEY FOR FIVE ELEMENTS, ABCDE

Column 1. Scores when the correct order is dbaec
Column 2. Scores when the correct order is beadc
Column 3. Scores when the correct order is bceda
Column 4. Scores when the correct order is ebacd

a Sets	1	2	3	4	b Sets	1	2	3	4	c Sets	1	2	3	4	d Sets	1	2	3	4
			abcd	set				baed	set				cabd	set				dabc	set
abcde	5	5	5	5	baede	6	6	6	6	cabde	3	3	5	3	dabee	8	4	4	2
abced	4	6	6	6	baced	5	7	7	7	cabed	2	4	6	4	dabec	9	5	3	3
abeed	5	7	5	7	baeed	6	8	6	8	caebd	1	3	5	5	daebc	8	4	2	4
aebed	4	6	4	8	beaed	5	9	7	9	ceabd	0	4	6	6	deabc	7	5	3	5
eabed	3	7	5	9	ebaed	4	8	6	10	ecabd	1	5	5	7	edabc	6	6	4	6
			abdc	set				bade	set				cadb	set				dacb	set
abdce	6	6	4	4	badce	7	7	5	5	cadbe	4	2	4	2	dacbe	7	3	3	1
abdec	7	7	3	5	badec	8	8	4	6	cadeb	3	1	3	3	daceb	6	2	2	2
abedc	6	8	4	6	baedc	7	9	5	7	caedb	2	2	4	4	daecb	7	3	1	3
aebdc	5	7	3	7	beadc	6	10	6	8	ceadb	1	3	5	5	deacb	6	4	2	4
eabdc	4	8	4	8	ebadc	5	9	5	9	ecadb	2	4	4	6	edacb	5	5	3	5
			acbd	set				bcad	set				cbad	set				dbac	set
acbde	4	4	4	4	bcade	5	5	7	5	cbade	4	4	6	4	dbace	9	5	5	3
acbed	3	5	5	5	bcaed	4	6	8	6	cbaed	3	5	7	5	dbaec	10	6	4	4
acebd	2	4	4	6	bcead	3	7	9	7	cbead	2	6	8	6	dbeac	9	7	5	5
aecbd	3	5	3	7	becad	4	8	8	8	cebad	1	5	7	7	debac	8	6	4	6
eacbd	2	6	4	8	ebcad	3	7	7	9	ecbad	2	6	6	8	edbac	7	7	5	7
			acdb	set				bcda	set				cbda	set				dbca	set
acdbe	5	3	3	3	bcdae	6	4	8	4	cbdae	5	3	7	3	dbcae	8	4	6	2
acdeb	4	2	2	4	bcdea	5	5	9	5	cbdea	4	4	8	4	dbcea	7	5	7	3
acedb	3	3	3	5	bceda	4	6	10	6	cbeda	3	5	9	5	dbeca	8	6	6	4
aecdb	4	4	2	6	becda	5	7	9	7	cebda	2	4	8	6	debca	7	5	5	5
eacdb	3	5	3	7	ebcda	4	6	8	8	ecbda	3	5	7	7	edbca	6	6	6	6
			adbc	set				bdac	set				cdab	set				dcab	set
adbce	7	5	3	3	bdace	8	6	6	4	cdabe	5	1	5	1	dcabe	6	2	4	0
adbec	8	6	2	4	bdaec	9	7	5	5	cdaeb	4	0	4	2	dcaeb	5	1	3	1
adebc	7	5	1	5	bdeac	8	8	6	6	cdeab	3	1	5	3	dceab	4	2	4	2
aedbc	6	6	2	6	bdeac	7	9	7	7	cedab	2	2	6	4	decab	5	3	3	3
eadbc	5	7	3	7	ebdac	6	8	6	8	ecdab	3	3	5	5	edcab	4	4	4	4
			adcb	set				bdca	set				cdba	set				dcba	set
adcbe	6	4	2	2	bdcae	7	5	7	3	cdbae	6	2	6	2	dcbae	7	3	5	1
adceb	5	3	1	3	bdcea	6	6	8	4	cdbea	5	3	7	3	dcbea	6	4	6	2
adecb	6	4	0	4	bdeca	7	7	7	5	cdeba	4	2	6	4	dceba	5	3	5	3
aedcb	5	5	1	5	bedca	6	8	8	6	cedba	3	3	7	5	decba	6	4	4	4
eadcb	4	6	2	6	ebdca	5	7	7	7	ecdba	4	4	6	6	edcba	5	5	5	5

by using the inverse in each case: ceabd, cdaeb, adecb, and dcabe. The score for each of the responses based on the inverse order is 10 minus the corresponding score given in Table II. Hence another table of scores can be constructed easily for four additional correct orders.

It may be protested that locating a single order among a list of 120 is too difficult and time-consuming. The arrangement of Table II however has been designed to make it possible, with a little practice, to locate any one of

the 120 orders almost instantly. The principle used is simple. Each of 24 sets of permutations of 4 elements, abcd, is shown. Then the fifth element "e" is permuted in order with each permutation of these sets. Furthermore, all the orders of 4 are arranged in 4 columns, the a, b, c, and d groups being classified according to the first letter of the permutation.

To use Table II in locating the score to assign to a particular response for a given correct order, begin by ignoring the "e" element in the response. Then, using the *first* letter in the remaining 4 elements, locate the column in which the order belongs: a, b, c or d. Then use the second and third letters to locate the specific set in the group of five. Finally use the position of the "e" element to locate the arrangement desired. What score should be given to the response "eadcb" when the correct order is "dbaec"? The correct order, dbaec, determines the column to work in; it is column 1 in each of the four major sets. Ignoring "c" in the response gives adcb. This is in the a set, and at the bottom of the column because the ab and ac sets precede the ad sets; the second letters in each permutation are also in alphabetical order down the page. Finally any permutation beginning with "e" is the last one in each set of 5; hence, eadcb is the last permutation in the first column, and its score is 4. A little practice makes it possible to locate any permutation almost instantly. In case of 5 elements, the scores range from 0 to 10 with an expected mean of 5.

Since the non-negative scores in Tables I and II have a perfect negative correlation with the corresponding number of inversions, the distribution of the latter may be applied to these scores in testing statistical hypotheses. In the previously cited example the mean non-negative score was 7.7; the null hypothesis tested was whether for 5 elements this sample value represented a likely significant departure from a mean of 5, the chance expectation value. The difference $7.7 - 5$ or 2.7 is identical to the difference, but with the opposite sign, that was obtained by using the inversions themselves. This relation will always exist.

It is to be observed that the expression for the sampling variance of a mean, $\frac{\sigma^2}{N}$ can be used for small samples instead of Student's t distribution *because the true population variance for the inversion distribution is known.* This true value, rather than the sample estimate, s^2, should be used in the expression $\frac{\sigma^2}{N}$.

SUBJECT PREFERENCE AS RELATED TO TEST INTELLIGENCE, INTENDED VOCATION, COLLEGE EXPECTATION AND RACE OF HIGH SCHOOL SENIORS IN HAWAII

T. M. LIVESAY
Dean, University of Hawaii

Editor's note: The author finds important differences in the intellectual levels and subject preferences of high school pupils. Students planning to enter college show preferences for the conventional high school subjects.

HIGH-SCHOOL students commonly show considerable variation in their preferences for the subjects of the curriculum, but as to the actual causes of such differences we have little dependable information. Aptitudes, intelligence and teacher personality have all been found to play a part, but in most instances the numbers investigated have been too small for entirely reliable conclusions. The present study, in one approach to the problem, deals with the relationship between subject preferences and test intelligence (as measured by the 1935 edition of the American Council Psychological Examination for High School Graduates and College Freshmen), intended future vocation, expectation as to college entrance and race of 2199 high-school seniors in the Territory of Hawaii. This represents a complete cross-section as all high schools were included.

Some thirty-five different subjects were mentioned, but in order to provide adequate numbers and reduce the total to a workable basis for comparative purposes it was necessary to resort to grouping under fewer headings. The six divisions of Commercial, Expression (agriculture, art, music, etc.), Language, Mathematics, Science and Social Studies were arbitrarily chosen as the most logical and best fitting categories.

The distribution of test scores under these divisions, and the measures of central tendency and dispersion are given in Table I. It is clear that Commercial and Expression subjects were most popular with these students, and Language and Mathematics least so. A hint as to the reason for the popularity of the first two is suggested by the means—lowest of the six. It is probably to be expected that students of less ability should favor these subjects, hence the larger numbers since there are so many more in the lower levels of ability as measured by the test. In fairness, however, it should be pointed out that quite a number of good ability expressed a preference for

these fields. This is proof, of course, that other factors besides intelligence enter into the liking or disliking of a subject at the secondary school level.

The comparisons in terms of significant differences are given in Table II. That there are real differences in intelligence, as measured by the Psychological Examination, is shown by column five, where, with three exceptions, the chances of a true difference are 100 in 100. The Critical Ratios are all considerably larger than required for a reliable difference with the exception of Language and Mathematics, Language and Science, and Science and Social Studies, and in these three comparisons they are large enough to yield a probability in excess of 90 in 100. The differences are in favor of the Commercial group as compared to the Expression group; Language as compared to Commercial, Expression, Science and Social Studies; Mathematics as compared to Commercial, Expression, Language, Science and Social

TABLE I

DISTRIBUTION OF SCORES BY FAVORITE SUBJECT

Score	Commercial	Expression Subjects	Language	Mathematics	Science	Social Studies
295–309			1	2	1	
280–294			2	1	2	3
265–279		1	3	3	3	
250–264	1		4	3	4	3
235–249	1	1	4	8	4	4
220–234	2	4	14	7	9	8
205–219	9	6	10	16	11	11
190–204	6	4	12	15	16	19
175–189	16	14	22	19	21	16
160–174	31	17	24	22	41	17
145–159	35	20	34	28	36	28
130–144	39	27	29	21	49	37
115–129	65	40	27	28	41	31
100–114	80	64	25	23	42	24
85–99	80	71	19	23	40	36
70–84	71	71	25	11	39	31
55–69	48	77	19	6	14	26
40–54	15	47	6	2	10	15
25–39	5	13		1		3
10–24	1	2				1
Total	505	479	280	239	383	313
Mean	110.83	97.64	141.45	150.43	134.69	127.94
Q1	82.10	66.25	100.60	110.92	97.28	85.94
Median	106.09	91.23	139.83	147.41	131.68	124.92
Q3	135.29	103.34	176.36	186.25	165.95	162.43
SD dis.	39.840	42.555	54.195	52.530	49.395	53.820
SD mean	1.773	1.944	3.239	3.398	2.524	3.042

TABLE II

FAVORITE SUBJECT COMPARISONS IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Combinations	Difference between means	S D of Difference	Critical Ratio D / SD diff.	Chances in 100 that the true difference is greater than zero	Difference in favor of
Commercial Expression	13. 19	2. 631	5. 01	100	Commercial
Commercial Language	30. 62	3. 693	8. 29	100	Language
Commercial Mathematics	39. 60	3. 833	10. 33	100	Mathematics
Commercial Science	23. 86	3. 084	7. 74	100	Science
Commercial Social Studies	17. 11	3. 521	4. 86	100	Social Studies
Expression Language	43. 81	3. 778	11. 59	100	Language
Expression Mathematics	52. 79	3. 915	13. 48	100	Mathematics
Expression Science	37. 05	3. 186	11. 63	100	Science
Expression Social Studies	30. 30	3. 610	8. 39	100	Social Studies
Language Mathematics	8. 98	4. 694	1. 91	97	Mathematics
Language Science	6. 76	4. 106	1. 65	94	Language
Language Social Studies	13. 51	4. 444	3. 04	100	Language
Mathematics Science	15. 74	4. 233	3. 72	100	Mathematics
Mathematics Social Studies	22. 49	4. 561	4. 93	100	Mathematics
Science Social Studies	6. 75	3. 953	1. 71	96	Science

TABLE III

PERCENTAGE OF MALES SELECTING DIFFERENT VOCATIONS ACCORDING TO FAVORITE SUBJECT

Future Vocation

Favorite subject	Agriculture		Business		Clerical		Professions		Semi-professions		Skilled trades		Total number of cases
	N	Per cent	N	Per cent	N	Per cent	N	Per cent	N	Per cent	N	Per cent	
Commercial	4	3.3	39	46.4	71	61.7	8	4.6	----	----	16	9.8	138
Expression subjects	43	35.2	7	8.3	8	7.0	12	6.9	9	23.7	69	42.3	148
Language	3	2.5	6	7.1	9	7.8	17	9.7	9	23.7	11	6.8	55
Mathematics	22	18.0	15	17.9	8	7.0	52	29.7	11	28.9	18	11.0	126
Science	32	26.2	3	3.6	6	5.2	62	35.4	4	10.5	30	18.4	137
Social Studies	18	14.8	14	16.7	13	11.3	24	13.7	5	13.2	19	11.7	93
Total	122	100.0	84	100.0	115	100.0	175	100.0	38	100.0	163	100.0	697

Studies; Science as compared to Commercial, Expression, and Social Studies; and Social Studies as compared to Commercial and Expression subjects.

In terms of the ability of the student selection, as indicated by the means, the six divisions rank in the following order: Mathematics, Language, Science, Social Studies, Commercial and Expression subjects. The relatively small standard deviations, ranging from 1.773 to 3.398, indicate that these means are fairly reliable.

That there is some correlation between favorite high-school subject and intended future vocation is illustrated in Table III which gives the percentage of males* choosing particular vocations according to subject preferences. Of the 122 who expected to enter Agriculture the favorite high school subjects were Expression (usually agriculture) and Science; for the Clerical group, Commercial subjects; for the Business group, Commercial subjects; for the Professions, Mathematics and Science; for the Semi-professions, Expression subjects, Language and Mathematics; and for the Skilled Trades, Expression subjects (usually some variety of shop course). Considerable scattering is evident, but on the whole there seems to have been a decided tendency for these high-school boys to choose vocations somewhat related to the type of high-school subject which they found most to their liking.

TABLE IV

COLLEGE EXPECTATION ACCORDING TO FAVORITE SUBJECTS

Favorite Subject	Expected to enter college		Did not expect to enter college	
	Male	Female	Male	Female
	Percent	Per cent	Per cent	Per cent
Commercial	19.6	11.4	80.4	88.6
Expression	21.8	22.0	78.2	78.0
Language	54.9	56.6	45.1	43.4
Mathematics	58.8	52.3	41.2	47.7
Science	55.1	53.5	44.9	46.5
Social Studies	48.4	52.0	51.6	48.0

The relation between college expectation and favorite high-school subject is shown in Table IV. As in the case of future vocation it is clear that subject preferences give some indication as to college plans. Of those expecting to enter college, both sexes clearly preferred the so-called solids,

* Not all of these students had made a vocational choice, but 697 of the males gave a definite response. The females were excluded, as their vocational choices tend to be much more restricted.

TABLE V

RACE AND SEX COMPARISONS BY FAVORITE SUBJECT

Race	Commercial		Expression		Language		Mathematics		Science		Social Studies	
	Per cent		Per cent		Per cent		Per cent		Per cent		Per cent	
	M.	F.	M.	F.	M.	F.	M.	F.	M.	F.	M.	F.
Asiatic-Hawaiian	22.9	26.2	22.9	23.8	8.6	14.3	11.4	4.8	22.8	11.9	11.4	19.0
Caucasian	4.8	13.9	7.7	15.7	16.3	30.4	29.8	3.5	19.2	20.8	22.2	15.7
Caucasian-Hawaiian	12.8	24.2	21.3	21.7	19.1	34.9	23.4	2.4	8.5	7.2	14.9	9.6
Chinese	21.5	24.4	16.3	26.3	9.9	13.1	16.8	7.5	22.7	13.1	12.8	15.6
Filipino*	9.1	-------	27.3	-------	18.1	-------	9.1	-------	9.1	-------	27.3	-------
Japanese	20.9	32.2	23.4	23.8	5.5	15.8	13.6	5.4	21.6	12.3	15.0	10.5
Korean	19.0	17.1	9.5	21.9	4.8	17.1	19.0	4.9	28.7	19.5	19.0	19.5
Portuguese	35.6	38.2	24.4	29.4	6.7	8.8	15.6	5.9	4.4	17.7	13.3	-------

* There were too few Filipino females for consideration.

with Mathematics leading for the males and Language for the females. The bulk of those preferring Commercial and Expression subjects did not expect to go on to college. This again ties up with intelligence, as the lowest mean scores were in these last two groups.

Race and sex comparisons as to favorite subjects are given in Table V. Some very evident contrasts are apparent between the groups as to preferences for certain types of high school subjects. The Asiatic–Hawaiian, Chinese, Japanese, and Portuguese show quite clear preferences for Commercial and Expression subjects; the Caucasian, Language, Mathematics and Science; the Caucasian–Hawaiian, Language, Expression subjects and Commercial; the Filipino, Expression subjects, and Social Studies; and the Korean, Commercial, Science, and Social Studies.

The sex preferences, except for a few variations, follow a fairly general pattern—the males favoring Mathematics, Science and Social Studies in larger proportions than the females, with the reverse for Commercial, Expression and Languages. The exceptions are the Asiatic–Hawaiian, Chinese, and Korean in Social Studies, the Caucasian and Portuguese in Science, and the Korean in Commercial.

SUMMARY AND CONCLUSIONS

1. This paper presents material concerning the relation of subject preferences to test intelligence, intended vocation, college expectation and race and sex for 2199 high-school seniors in Hawaii.

2. There are significant differences in test intelligence in all but three of fifteen comparisons of subject preferences.

3. There is evidence of a tendency for high-school students to choose future occupations somewhat related to the favorite subject.

4. Of those expecting to enter college, the larger proportions preferred the so-called solids—Language, Mathematics, Science, and Social Studies.

5. There is evidence of race and sex differences in subject preferences.

SOME DIFFERENTIAL EFFECTS OF MOTIVATION UPON ACHIEVEMENT AND INSIGHT IN COLLEGE PHYSICS[1]

D. D. FEDER
University of Illinois
Lieutenant Commander, Bureau of Naval Personnel, Training Division, U. S. N. R.

and

M. ERIK WRIGHT
Ohio State University

Editor's note: There are apparently a large number of factors conditioning success in any area. The authors present data to show that the motivational factor is important in the study of physics.

IN AN educational system which emphasizes elective choice of subjects, it is generally assumed that every student who enrolls in a class does so with a definite purpose in mind, and with a reason whose validity will supply him with adequate motivation. This is a report of an effort to test the validity of the foregoing assumptions. The subjects were students in a first year course in college physics. The course was purely elective for some, but required for others because of the specific curriculum in which they were enrolled. The study was concerned with two major problems: (1) The degree of insight shown by the students as to the possible contributions of physics to their future college work, and how the year's work could be applied to everyday life situations; and (2) the effect of motivation (as evidenced by their reasons for taking the course) upon their achievement.

One hundred forty-two students, taught by the same instructor, acted as subjects. Scores on a series of general ability tests, a physics aptitude test, and a series of objective tests of achievement in college physics were available. Just before the end of the second semester each student was requested to fill out a short questionnaire which dealt with his reasons for taking the course, and asked for a free report of his opinion as to the possible relationship of physics to his future work in college and to everyday life situations. The key questions with which the present study is concerned were: "Will physics aid you in your future college work?" and "Can the physics you have learned to date be applied in everyday life situations?" The students were

[1] The assistance of Professor C. J. Lapp, Department of Physics, and Dr. John W. Hancock, formerly Research Assistant, Department of Psychology, University of Iowa, is gratefully acknowledged.

instructed to respond to each question with pertinent examples. The questions and instructions were so worded as to eliminate the possibility of the student's feeling that his responses might affect his grade in the course. Furthermore, the students were assured their responses would not be seen by the instructor until the semester was completed and grades reported. Evaluation of the responses to these two questions was made to determine the possibility of differentiating degree of insight and its possible relation to performance in the course. In order to balance the possible bias of a departmental point of view, two members of the physics department and the authors rated the responses in accordance with the following scale and directions:

By "insight of the student" we mean the quality of the student's understanding of the role of first year college physics, as expressed in the functional relationship of college physics to his academic life and to his everyday life out of school. Assign to each student's responses the numerical value corresponding to the statement which most nearly describes the quality of the responses.

1. Deep insight into possible functions and applications of physics.
2. Recognition of the importance of physics as a science and as related to other fields.
3. Belief that physics is important as a science; "pat answer" type of attitude.
4. Belief that physics has little importance except as a prerequisite course.
5. Belief that physics is unnecessary for understanding life phenomena.

Because of the generally high agreement among raters, the four ratings given each student on each question were averaged, and the averages for the two questions were combined into a single score. Table I shows the correlations obtained among the various raters in terms of this composite average score. As might be expected, the highest agreement was obtained between the ratings of the judges from the same department. Thus the physicists agreed with each other slightly better than did the psychologists, and markedly better than either physicist agreed with either psychologist.

The coefficient of .85 $\pm$ .02 in Table II is spuriously high, of course, due to the self-correlation present. Despite the relatively high correlation ($r = .73 \pm .04$) between scores on the college aptitude test and the physics aptitude test, the latter is markedly superior in the prediction of actual

TABLE I

AGREEMENT OF RATINGS BY VARIOUS RATERS

Raters	r	PEr
Ph^1 vs. Ph^2	.89	.01
Ps^1 vs. Ps^2	.82	.02
Ph^1 vs. Ps^1	.71	.03
Ph^1 vs. Ps^2	.75	.03
Ph^2 vs. Ps^1	.63	.03
Ph^2 vs. Ps^2	.65	.03

Ph = Physicist
Ps = Psychologist

TABLE II

INTERCORRELATIONS OF APTITUDE, INSIGHT, AND ACHIEVEMENT SCORES

	1	2	3	4	5	6	8	9
2	.85 ± .02							
3	.87 ± .02	.61 ± .04						
4	.47 ± .06	.42 ± .06	.41 ± .06					
5	.23 ± .06	.56 ± .04	.34 ± .06	.73 ± .04		.68 ± .03		
7	.30 ± .06	.27 ± .06	.23 ± .06	.24 ± .07	.40 ± .05	.24 ± .05	.48 ± .02	.15 ± .06
8	.31 ± .06	.22 ± .06	.26 ± .06	.29 ± .07	.30 ± .05	.27 ± .05		.03 ± .06

KEY

1. Total Grade Point Average*
2. Science Courses Grade Point Average
3. Non-Science Courses Grade Point Average
4. Freshman College Aptitude Examinations
5. Physics Aptitude Test—Iowa Placement Examinations
6. Physics Grade Average
7. Average Insight Score—"Future College Work"
8. Average Insight Score—"Everyday Situations"
9. Total Number Semester Hours of Science Courses Taken

*A = 4 points; B = 3 points; C = 2 points; D = 1 point; E = 0 points

achievement in physics ($r = .68 \pm .03$ as compared with $.42 \pm .06$), supporting previous studies indicating the superiority of a specific aptitude test for predicting specific subject matter achievement.[2]

Most noteworthy of the relationships between quality of insight and the other measured factors is the coefficient of $.40 \pm .05$ between physics aptitude scores and scores derived from responses dealing with the relation of physics to other college work. The relatively low correlations ($.24 \pm .05$ and $.27 \pm .05$) between the scores on the insight items and other measured

[2] Feder, D. D. "An Evaluation of Some Problems in the Prediction of Achievement at the College Level," *Journal of Educational Psychology* (November, 1935).

achievement, suggest the possibility of measuring a factor in physics achievement heretofore considered to be a part of routine problem-solving, but which actually appears to be relatively independent of other aspects of achievement.

As indicated previously, it is possible to discover a wide variety of reasons for students' enrolling in first-year college physics. In terms of their offered reasons, the students were classified as follows:

Group I: Elective. Students were majoring in Journalism, Art, Commerce, Psychology, English, etc., and were taking the course in order to fulfill degree requirements in science, or for personal reasons.

Group II: Required course. Students were majoring in a science field such as Engineering, Geology, Pharmacy, Chemistry, etc.

Group III: Pre-requisite course. Students were in the pre-medical and pre-dental curricula, or in some related field for which the course was prescribed as training necessary for future work in the chosen vocational field.

From the data in Table III it appears that the Group III students were somewhat superior to the other two groups in physics aptitude. This superiority is further manifested in average scores earned on the year's achievement measures. Although the Group I students as a whole were slightly superior to those in Group II, their average achievement in physics was actually slightly lower than that of Group II. None of these differences is statistically significant. The slight superiority of the Group III students is maintained for the total grade average and is even more marked for the total of science subjects. However, their grades in non-science subjects drop markedly below the same grades for Group I. The non-science grades of Group III are higher than those of Group II probably because of the actual superiority in scholastic ability of the former group.

Although Group I students showed less variability in physics aptitude, they showed greater variability in the various achievement areas studied than did the other groups. Since, by definition, a low score indicates superior insight, on both items Group III is superior to the other two groups.

Groups II and III obviously have more in common as far as goals in taking the course are concerned, because their ultimate vocational intentions involve scientific work. However, since there was a marked ability difference between the two groups (Table III), the question then arose as to how these

TABLE III

COMPARISON OF THE THREE MOTIVATIONAL GROUPS IN TERMS OF ABILITY, ACHIEVEMENT, AND INSIGHT

	Group I		Group II		Group III	
Item	A. M.	S. D.	A. M.	S. D.	A. M.	S. D.
1	2.39	.80	2.36	.59	2.55	.56
2	2.05	.79	2.33	.54	2.53	.66
3	2.59	.77	2.30	.61	2.49	.61
4	560		527		559	
5	116	18.40	114	24.10	126	22.80
6	17.78	4.29	18.57	3.95	19.37	3.59
7	3.27	1.21	3.30	.95	2.77	1.19
8	3.06	.94	3.01	.98	2.77	.68
N	24		65		53	

KEY

1. Total Grade Point Average*
2. Science Courses Grade Point Average
3. Non-Science Courses Grade Point Average.
4. Freshman College Aptitude Examinations
5. Physics Aptitude Test—Iowa Placement Examinations
6. Physics Grade Average
7. Average Insight Score—"Future College Work"
8. Average Insight Score—"Everyday Situations"

N = Number of Cases

*A = 4 points; B = 3 points; C = 2 points; D = 1 point; E = 0 points.

two groups would compare in terms of the insight items if the ability factor were held constant. To answer this question, 36 students from Group II were matched with a like number from Group III in terms of their scores on the Physics Aptitude test. Table IV gives the evaluation of the mean differences in achievement and insight of these matched groups. Although the difference in physics grades earned by the two groups is not a reliable one, the difference for ratings given on Insight Item III is statistically significant, and that for Item IV closely approaches statistical significance. In the matched group analysis, both the insight differences favor those students who were science majors, whereas pre-medical and pre-dental students, who might be expected to be more on the lookout for applications of the subject matter, actually found fewer of them than did the "pure" science students.

SUMMARY

This study is a report of an effort to devise a means of evaluating effects of differential motivation of students upon their achievement in college physics. In addition to the usual criteria of achievement, an effort was made to evaluate insight into the subject matter as revealed by the student's

TABLE IV

EVALUATION OF THE MEAN DIFFERENCES IN ACHIEVEMENT AND INSIGHT OF THE MATCHED GROUPS II AND III

	Diff.	P. E. Diff.*	C. R.	Favor of
Physics Grades	.19	.38	.50	Group III
Insight Item III	.64	.11	5.82	Group II
Insight Item IV	.36	.12	3.00	Group II

* The formula for "matched" groups described by E. F. Lindquist, Journal of Educational Psychology, March, 1931 was employed.

ability to make applications of it to everyday life and to his future college work, through specific examples. These examples were evaluated by two members of the physics department, as well as by the authors. On the basis of the obtained data, the following generalizations are offered:

1. In evaluating the significance of the student's responses, the physicists showed slightly higher mutual agreement than did the psychologists, although all the intercorrelations were high enough to permit the averaging of the four ratings.

2. A specific aptitude test, in this case the Physics Aptitude test of the Iowa Placement Examination Series, had higher predictive validity for achievement than did a general ability test battery. None of the predictive measures had significant validity for predicting ratings on the insight items.

3. Ratings on the insight items bear a low positive relationship to the measured aspects of physics achievement.

4. Despite a superiority in ability, those students who were taking the course as part of the science requirement in the liberal arts curriculum were slightly poorer in achievement than the others.

5. Students who were taking the course as part of prerequisites for entrance into the colleges of medicine and dentistry showed a slight superiority in achievement. This superiority persisted even when these students were matched on the basis of physics aptitude with a group of science majors. The difference was not statistically significant.

6. In general, the group who were taking the subject purely as a requirement, where no vocational or major motivation was present, exhibited a greater disparity between ability and achievement than did either of the other two groups.

7. On both the items purporting to measure insight, the pre-medical and pre-dentistry and liberal arts students obtained poorer scores than the "pure" science students. When the pre-medical and pre-dentistry group was matched in terms of ability with the group composed of science majors, these differences became even more pronounced in favor of the latter group.

The findings of this study as regards "insight" must be regarded as preliminary and tentative. However, for the instructor who emphasizes functional applications of instructional material, the technique may offer a crude but meaningful measure of factors not now being sampled by usual forms of measurement.

STUDENTS HELP SET UP CRITERIA TO AID IN DECIDING WHAT TO STUDY

RAY H. SIMPSON
Assistant Professor of Psychology
University of Alabama

THE material in this article represents a summary of some of the ideas drawn up by the writer in cooperation with students in several of his classes in History of Education and in Educational Psychology. Since other teachers, regardless of field or level of instruction, and other students also face the problem of trying intelligently to decide what to study, this article is being published. The writer wishes to emphasize that little value is likely to be derived from these criteria by teacher or student if they are read and accepted uncritically. Rather it is strongly felt that it is vitally important that each teacher and his students think through and work out their own set of criteria for deciding what problems they will study at a particular time. It is hoped that the points suggested in this article will direct thought and attention to some of the problems involved.

In the following paragraphs some *suggestive* answers are given to the following questions: (1) Why is there a need for teachers and students in all fields and at all levels to give attention to the question of what is *best* to study at a particular time and in a particular place?, (2) What are some criteria which may be worthwhile to apply when a student is trying to decide with the help of the teacher what is best for him to study?, and (3) What are suggestions and cautions which teachers should keep in mind while attempting to guide the learner in developing the ability to select wisely what to study?

REASONS FOR GIVING ATTENTION TO THE DEVELOPMENT OF THE ABILITY TO DECIDE WISELY WHAT TO STUDY

1. The deciding of what to study constitutes the first main step in efficient learning. When this decision is made wholly or almost wholly by the teacher the learner is getting little or no experience or practice in taking this initial step in getting the learning process under way. It is now recognized by psychologists and many others that unless the learner has an intrinsic purpose in studying his work, while it may be done as directed, it is

likely to have as little effect on his behavior as the work of the boy who made the mistake of writing "I have went." He was kept in after school and directed to write, "I have gone" one hundred times. He stayed in, wrote "I have gone" one hundred times, and then left this note to the teacher who was out of the room at the time: "Teacher, I done what you told me to do and I have went home." Thus, it too frequently is if the learner has not been consulted when the decision is made as to what, when, and how he shall study.

2. With many school and teachers attempting to put decreasing emphasis upon marks, credits, and simply sitting in class, other and more basic types of motivation need to be brought into play if better learning is to result. One of the best motivators unquestionably is to have the learner participate in the decision as to what he shall attempt to study. Certainly this does not mean that he shall or should be permitted to decide alone what he shall study at least until he has much guided experience. It does mean that instead of the teacher dictatorially making this decision the learner plays a gradually increasing part in making decisions which affect his own welfare.

3. The increasing recognition on the part of wide-awake teachers (1) that our present educational setup is not adequately meeting individual educational needs coupled with the realization (2) that it is impossible to meet these needs as long as teachers reserve almost exclusively for themselves the responsibility to make all assignments and evaluations for every learner, decide what it is "best" for every learner to study, select the materials for every learner to study, and even decide almost exactly when and how the learner should study what the teacher has picked out for him. *Not only do time limitations make this an impossible job for the teacher but such procedure also takes away from the learner the chance to develop in the ability to carry on these activities wisely.* Mass education can take care of individual needs if teachers will help the student develop the ability to decide intelligently: (1) What it is best for him to study, (2) What resources should be used in studying these problems, and (3) How to evaluate what he has done so that he will do a better job of learning in the future.

4. President Seymour of Yale in his Annual Report published October, 1941, says: "The educated man is the one who has learned to educate himself." The truth and significance of this statement is increasingly becoming recognized by wide-awake teachers and administrators. If the activities carried

on during formal schooling do not lead to organized education in the several areas after school then the education has been distinctly limited in value. To make the years of formal schooling just the beginning of the learner's education rather than both the beginning and the end, it is vitally important that the learner have much practice in selecting the problems it is best for him to work on in order to have a happier and better life.

5. Mere recognition on the part of the teacher of the need for aiding students in learning how to decide wisely what to study is far from sufficient insurance that a teacher will do a good job of this. To lead learners properly in developing this ability is an art which requires years of study on the part of the teacher. Most teachers now trying to teach learners the ability being discussed in this article are frequently frank to admit they are rather baffled. One teacher expressed it this way: "I know each learner should gain more independence in wisely deciding what he should study, but I don't know how to go about it."

6. If there is to be a gradual transition from the present situation where teachers assume almost all the responsibility for deciding what should be studied to a situation where the learners are given increasing responsibility for such decisions, the teacher must work out with learners the best way of doing this in a particular situation. Unless the pupils are properly approached they are likely to rapidly become disillusioned and feel that it is not worth the trouble.

7. The spreading of the "core" curriculum movement makes it imperative that the teacher and learner give additional consideration to reliable bases upon which to decide what it is best to study.

8. Functional democracy in education implies that the learner shall share in the making of decisions to the extent to which he is capable of sharing and in which he needs practice. This is more vital to most learners than just reading *about* democracy.

SUGGESTED CRITERIA TO BE APPLIED IN THE SELECTION OF AN APPROPRIATE PROBLEM FOR STUDY

As was indicated in the first paragraph teachers or students are *not* encouraged to simply take this list and apply it to possible problems for study. Rather teachers and prospective teachers *are* encouraged to study these suggested criteria and then to try to encourage students to develop a set of

standards consonant with their degree of ability and maturity. It should be remembered that if criteria are to be basically meaningful, those who use them should have a share in developing them. Individuals in groups with whom the writer has worked have found the following questions helpful in intelligently arriving at a sound problem for study.

1. Can I see how working on this problem is likely to help me meet one of my present needs or prepare me better to meet prospective needs in one or more of the following:

a. An occupational situation? This should lead to (1) a consideration of several types of work in which the learner is most likely to engage in view of his intelligence, financial backing, occupational desire, etc., and (2) a consideration of the types of problems he is likely to face in such occupations. Obviously, it is impossible to predict with certainty what the learner will actually be doing ten years from now but the approach suggested is likely to lead him to work harder and make him more thoughtful about his future problems.

b. A social situation? For example, what types of speaking knowledges and skills are likely to be required for success or at least decided assets in a conversational group? In a discussion group? In an office situation? In a courting situation?

c. A recreational or leisure time situation? For instance the ability to engage in certain types of games or other activities of a group or an individual nature is likely to be an asset. What are these activities and how should I go about developing them?

d. A parental situation? What types of things do happy parents know? What skills related to raising children take some time and guidance to develop properly?

e. A situation involving my work in making my community a better place in which to live? For example, do I have the training necessary to present a playground problem to leaders in the community or do I know how to go about voting intelligently? If not, would studying this problem probably help me to prepare for this type of situations so important in a democratic community?

If the answers to any or all of the questions given above is, "yes", then the questions given in succeeding paragraphs should be considered.

2. Is this a problem which I have already studied in an organized fashion? If so, are there other problems it would likely be more valuable for me to study? Do I need additional work experience before giving more time to an intensive study of this problem?

3. What might be the advantages of studying the problem at a later date? What would probably be the disadvantages of postponing the study of the problem?

4. Is my realization of the need for a study of this problem great enough to keep me working on it for a reasonable length of time? If not, perhaps it would be well to search farther for a more appropriate problem.

5. Does it seem likely that I will learn better how to *select* or solve problems through a serious consideration of this problem now proposed for study?

6. Is it likely that there will be a sufficient number of available sources, human and otherwise, to facilitate and justify my study of this problem at this time and in this place? (The possibility of using nonprinted sources of help such as interviews with parents, other learners, teachers, observations, letters, and workers in various occupational fields should constantly be kept in mind. Also, magazine, newspaper, and pamphlet sources are often more usable than book sources. Always the advisability of constructing and using informal experimentation should be kept in mind.)

7. In view of limited experience in problem solving, is the problem under consideration simple enough for me to attack without getting swamped in complex ramifications? For example, the study of "What is a desirable method of national price control?" is likely to be less appropriate from this standpoint for most students than is a study of "How can I learn to buy more wisely?" or "How should I make and use a budget?"

8. Am I seeing this problem sufficiently in relation to larger and more complex problems of which it is a part? For instance, if the learner is considering the study of "How can I become a wiser buyer and consumer?" does he consider the relationship of this problem to such organizations as "Consumers Research", "Consumers Union", and governmental control of buying and selling?

9. Could work on this problem be related to work being done in other courses? Is this desirable or undesirable? Why do I think so?

10. Is the problem specific enough so that I am likely to get practical answers to my questions which answers will actually improve my way of living and tend to increase my happiness? The mere "Study of Health" is not as likely to meet this criterion as "What can and should I do to improve my own and others health in this community?"

11. In what way might work on this problem be related to the fundamental skills generally considered a part of the work of a course of this sort? For instance, could the problem of "How can we go about actually raising a school garden?" be related or be part of the work in an English course? (One answer might be that this would provide many opportunities for practical and functional writing, speaking, and reading under guidance which might be more useful than formalized English exercises.)

12. Should my likely needs be considered more important than the traditional concept of what should be done in a course? For instance, if the traditional way to start a ninth grade class in American History is to give everybody a ninth grade text and start them studying the first chapter, should this be done in preference to a study of keeping an efficient personal educational history if the latter seems to be more needed by me as a learner? What reasons might you advance for your answer?

13. How are parents, school officials, and others in the community likely to regard my study of a problem of this sort? If they are likely to oppose it but otherwise it seems like a "good" problem to study how might I best approach this difficulty? For instance, a study of "What are the probable advantages and disadvantages of strikes and how should I regard a strike in the local mill?" might be thought undesirable by certain groups in the community. In other communities it might be negro problems or problems related to evolution. Public opinion should be considered when decisions are being made with respect to what to study.

14. With the help of the teacher have I sufficiently investigated or am I going to investigate several problems before starting to work on any single problem? To work on the *best* problem implies that more than one problem has been considered and then one of the possibilities has been picked out.

SUGGESTIONS AND CAUTIONS TO TEACHERS WHO SERIOUSLY ATTEMPT TO HELP THE LEARNER DEVELOP THE ABILITY TO DECIDE WISELY WHAT TO STUDY

1. A study of this type should probably first center upon ourselves as teachers. Each teacher should perhaps first ask himself the question: "How can I more effectively further my own education by studying my personal and professional problems and then systematically do something about

them?" Obviously, it will be extremely difficult for a teacher to teach a skill in which he himself is not rather skillful.

2. Remember that what you are trying to teach is a very complex process which we should not expect to have satisfactorily developed in a week, a month, or even a year. It will be a hard pull if done in the proper way.

3. Keep in mind that old ideas, attitudes, and habits relating to the question of what constitutes education and schooling will die slowly if at all. These ideas, attitudes, and habits have been tied up in the case of most learners with a teacher dictated type of education which tends to stop when schools stops since the learner has been given little or no opportunity to go ahead and assume much responsibility in getting himself educated.

4. Recognize the fact that learners are still being strongly influenced by teachers and classmates who assume that the primary job of the teacher is to decide what cans of information are to be memorized, open those cans, and, then dish out the facts to be taken in by the learner with little or no thought as to when if ever he is likely to need them. Obviously under these circumstances learner initiative and responsibility tend to die a lingering death.

5. Accept the fact that the learners will make many mistakes of judgment when he is given a chance to help decide what he shall study but also remember that it is only when there are some mistakes that we have learning. The only student who makes no mistakes in deciding what he shall study is the one who never has a chance to make any decisions of this sort. Great amounts of patience and tolerance are vitally necessary if functional learning is to proceed rapidly.

6. Take care not to encourage the learner to work *too exclusively* on problems of great interest but devoid of values aside from the recreational one involved. Too much time spent on units such as one on Mexico—houses, roads, sanitation, education—may lead to *an escape* from the immediate problems related to providing better living conditions right in the local community.

7. Do not be too upset if the learners are somewhat bewildered and confused by the use of new methods. A certain amount of "puzzlement" is a necessary concomitant to effective learning.

8. If you want the *easy way* to teach a class you as teacher decide what is to be studied, when it is to be studied, how it is to be studied—give the

learner no chance to participate in these decisions. That teacher follows no easy road who *tries to train a learner to decide wisely what he should study,* when he should study it, and how it is to be studied. Dictatorial methods are far easier than developing functional democracy in the school. However in this picture it should be remembered that the satisfaction which comes with knowing learners are growing in the ability to educate themselves is pay fit for an educational king. The modern teacher may be busier but she also can be happier than the old-time dictatorial teacher.

9. Remember the teacher has a new role to play if the idea is accepted that the learner should have a share in deciding what he shall study. Teacher does not have all the answers up her sleeve. Teacher serves as a guide rather than a dictator of study. Teacher may not know information but can suggest to learner places where he can likely find it.

THE COURSE CONTENT IN COMMERCIAL MATHEMATICS

WALTER F. CASSIDY
St. John's University, Brooklyn

Editor's note: Much has been done in recent years to bring the curriculum into closer relation with the life needs of pupils and grownups. The author reports a study of the arithmetical needs of business and commercial workers.

WITHIN the past quarter-century there has been much controversy over what should be included in a course in commercial mathematics. In 1934, a large group of office managers rated such a course second among practical commercial subjects.[1] Yet, in many places, a course of this kind has fallen into disrepute, with the result that fewer schools are offering it. The first term of a course in commercial mathematics usually consists in repeating the elementary school work in addition, subtraction, multiplication, division, and common and decimal fractions. In such repetition the student's interest is seldom challenged, and often the course is a waste of time.[2]

The content of the commercial mathematics course, as now taught in most schools, has no scientific reason for existence, since it is not known, as yet, how closely this subject matter satisfies the needs of the business world. Moreover, thinkers in the field suggest that the informational aim should be given greater recognition than the operational, because of the prevalence of calculating machines in business offices. Furthermore, business men are justly annoyed by the school's failure to teach a mathematics vocabulary, since the chief difficulty of the business apprentice seems to be his lack of understanding of language concepts behind problem solving. The pupil entering business with a precise and adequate terminology, thoroughly understood, would have his apprenticeship period shortened considerably. Thus, the controversy concerning the content of the commercial mathematics course is still on. The writer of this investigation has attempted to validate some basic mathematics items for the commercial mathematics curriculum. This justification, or validation, has been made by subjecting commercial mathematical items to specific tests or criteria for social and vocational usefulness. By usefulness is meant fitness for some desirable, practical purpose. Social

[1] C. L. Bailey, "A Radio Program in Commercial Education," *Journal of Business Education*, X (November, 1934), p. 7.
[2] H. E. Stelson, "Proposed Course in Business Arithmetic," *School Science and Mathematics*, XXXIV (January, 1934), p. 91.

usefulness, in this investigation, means the preparation of the individual, through a course in commercial mathematics, for effective social contacts and relationships with his associates in his daily community life. Vocational usefulness, in this investigation, means the preparation of the pupil, through a course in commercial mathematics, for efficient business life.

The two major bases for this validation, therefore, were social and vocational utility. Specifically, the writer has set up 6 separate and independent criteria to determine "usefulness." One of the 6 criteria is concerned with validation for social utility, and the remaining five criteria are concerned with validation for vocational utility. Thus, for any mathematical item to be finally validated, it must have satisfied either the one criterion for social utility, or two or the five criteria for vocational utility.

Briefly, these criteria are as follows:

1. The first criterion comes from an analysis of seven theses which treat of the social usefulness of some mathematical items. From these seven theses, the writer has set up a list of items which were found to have occurred at least once in four or more of the seven. For an item to satisfy the criterion for social utility, it must have been found already validated in at least four of seven theses.

While it is true, at first glance, that only one criterion has been set up for the purpose of validating items by the test of social utility, yet, it should be held in mind that this one criterion is a composite of several other separate criteria established earlier through the work of other investigators in the field.

2. The second criterion, a test of vocational utility, flows from an analysis of seven high school and college accounting and bookkeeping text-books in which certain commercial mathematical items were more or less commonly found. From these seven text-books, the writer has made a list of mathematical items which were found to have occurred at least once in four or more of seven text-books. Therefore, for an item to satisfy this second criterion, it must have been found at least once in four or more of seven high-school and college accounting and bookkeeping text-books.

3. The third criterion, a test of vocational utility, is derived from an analysis of 155 pieces of sales promotion literature of six nationally known business machine companies, in which certain commercial mathematical items were found. From these 155 pieces of sales promotion literature, the writer

has established a list of items which were found at least once in the sales literature of three or more of six business machine companies. Therefore, for an item to satisfy this third criterion, it must have been found at least once in the sales promotion literature of three or more of the six business machine companies.

4. The fourth criterion, a test of vocational utility, arises from an analysis of 17 New York City Civil Service Examinations for accountant, bookkeeper, and clerk, in which certain commercial mathematical items are found.

From these 17 civil service examinations, the writer has formed a list of items which were found to have occurred at least once in nine or more of 17 examinations. Therefore, for an item to satisfy this fourth criterion it must have been found at least once in nine or more of the 17 civil service examinations.

5. The fifth criterion, a test of vocational utility, is set up from an analysis of one year's issue of each of six national trade journals in which some commercial mathematical items are found. From these six national trade journals, the writer listed items which occurred at least once in half or more of the number of copies in one year's issue of a single trade journal. For an item to satisfy this fifth criterion it must have been found at least once in half or more of the number of copies in one year's issue of a single trade journal.

6. The sixth criterion, a test of vocational utility, grows out of an analysis of the answers to a check list of commercial mathematical items submitted to 50 certified public accountants. The writer has set up a list of items which were found in 26 or more of 50 answers of these accountants. Therefore, for an item to satisfy this sixth criterion, it must have been found in 26 or more of 50 answers of certified public accountants.

From the analysis of these six foregoing sources, the writer has established one criterion for determining the social usefulness of some items in commercial mathematics (which criterion is, of course, a composite of several other criteria), and five criteria for determining the vocational usefulness of these same items. For an item to be validated under the criterion for social utility it must have been found in at least four of seven theses. Yet, even though an item, finally, has not met the criterion for social utility, it may, however, be validated by satisfying two of the five criteria for vocational

utility. Thus, even though an item fails to satisfy the one criterion for social utility, it might, quite possibly, satisfy the criteria for vocational utility. If it does so, the item is validated for the purposes of this investigation.

The scope of this study is limited to a validation of some basic mathematical items for the commercial mathematics curriculum. In time, it is limited to the period from 1928 to 1939 because: (1) this period is the one in which most investigations have taken place; and (2) in this period, business machines became common in office practice and, hence, business needs and business practice caused educational requirements to change considerably. This study is not concerned with any of the operational skills, but is limited solely to informational mathematics for the commercial mathematics curriculum.

This present study is of value to the following: (1) Those who teach commercial mathematics and are interested in a validation of certain items in their subject. (2) Curriculum builders who seek to determine the basic content of a course in commercial mathematics or elementary business training. Elementary business training is a combination of mathematics and economics. (3) Those educators who wish to eliminate much of the operational arithmetic in the course of study, and replace it with informational arithmetic. (4) Individuals who are interested in civil service preparation, and who wish to secure vocational information from the analysis of the mathematical content of these examinations. (5) Those educational agencies which offer courses in the operation of business machines, and desire to have readily at hand the basic content in commercial mathematics which the student should be given.

The following steps comprised the procedure: (1) the preparation of a check list of mathematical items taken from 10 text-books in commercial mathematics; (2) the analysis, for mathematical items, of seven theses, seven text-books, 17 civil service examinations, 155 pieces of sales promotion literature of six nationally known business machine companies, one year's issue of each of six national trade journals, and the answers of 50 certified public accountants; (3) the preparation of the necessary tables; (4) the organization and presentation of the data under three categories, namely: (a) Accounting and Bookkeeping Items, (b) Banking and Investment Items, and (c) Graphic Items; (5) the enumeration of items found in each of the six sources; (6) the determination of the composite rank of each item from

the individual ranks which the item held in each of the six sources; (7) the validation of the mathematical items in the check list by the application of the criterion for social utility; (8) the validation of the mathematical items in the check list by the application of the five criteria for vocational utility; (9) the determination of the degrees of validation of the items in the check list, when the criteria are increased beyond two; (10) the summarizing of the findings of the study; and (11) the drawing of conclusions from the findings.

Budgeting and simple interest were the only items that met the one criterion for social utility and the five criteria for vocational utility.

Sixteen items met the one criterion for social utility and four of the five criteria for vocational utility: annuities; assessed valuation on property; assets of business; depreciation of assets; cash; credit; debit; income; compound interest; life insurance; installment buying; inventory taking; mortgage; tax rate; income taxes; and real estate taxes.

Three items met the one criterion for social utility and three of the five criteria for vocational utility: checking accounts, bank discount, and promissory notes.

Three items met the one criterion for social utility and two of the five criteria for vocational utility: building and loan associations, trade acceptance, and foreign money.

Twenty-eight items met only four of the five criteria for vocational utility: accounts payable; accounts receivable; balance sheet; bank balances; commission; cost; cost of bonds; debit; cash discount; premium on bonds; trade discount; cost of labor; productive labor; liabilities; fire insurance; operating expenses; overhead expenses; parcel post; payroll distribution; statement of profit and loss; gross profits; net profits; surplus profits; sales records; dividends from stock; market value of stock; par value of stock; and tax on stock.

Twenty-seven items met only three of the five criteria for vocational utility: depletion of assets; bankruptcy; bonus; credit rating; freight rate; savings bank accounts; discount on bonds; yield of bonds; real estate brokerage; stock brokerage; unemployment insurance; net worth; discounting of notes; interest-bearing notes; maturity of notes; partial payments; nonproductive labor; organization of corporations; distribution of profit and loss in partnerships; undivided profits; cost of transportation; reconciliation

of bank statements; common stock; preferred stock; stock without par value; stock exchange; and hour rate system of wages.

Twenty-three items met only two of the five criteria for vocational utility: amortization of bonds; bank draft; commercial draft; sight draft; time draft; accident insurance; automobile insurance; currency denominations for payroll; deficit; pensions; reserves; sales taxes; cash surrender value of life insurance; health insurance; money order; divided bar graph; horizontal bar graph; vertical bar graph; single curve graph; several curves plotted on one graph; broken line graph; wages for overtime work; and wages for piece work.

Seven items met only one of the five criteria for vocational utility: equations of accounts; duty; excise taxes; circle graph; pictograph; scale drawing; and square root.

The following conclusions were drawn from the study: (1) It is evident from degrees of validation that if the criteria for validation by vocational utility were increased beyond two, many items, which authorities in the field consider necessary, would have to be omitted from the commercial mathematics curriculum; (2) A course in commercial mathematics can be built around the validated items with some degree of scientific accuracy; (3) Since every item which satisfied the criterion for social utility was found to have satisfied also at least two of the five criteria for vocational utility, it would seem that it is extremely difficult to differentiate, in the final analysis, between social and vocational utility in the validation of items for the commercial mathematics curriculum; (4) The items simple interest and budgeting, which satisfied the majority test of the civil service examinations, are two of the most essential items in commercial mathematics. These items are important because of their social and vocational usefulness, as shown by their meeting the one criterion for social utility, and the entire five criteria for vocational utility (text-books, business machine literature, trade journals, civil service examinations, and the answers of accountants); (5) Many of the mathematical items found in the trade journals such as, pensions, unemployment insurance, etc., reflect the emphasis on social and vocational security in contemporary life. Thus, such items are important for the commercial mathematics curriculum.

AN ANALYSIS OF TECHNIQUES FOR DIAGNOSTIC VOCABULARY TESTING

LEE J. CRONBACH
State College of Washington

Editor's note: The author reviews the literature on diagnostic vocabulary testing and finds that diagnostic testing requires more valid instruments of measurement.

CONSIDERABLE discussion in the past decade has summarized the difficulties of vocabulary research and clarified the problems to be solved so that measurement of the pupil's word knowledge is fairly dependable. This measurement may seek to determine approximately how many words the pupil knows, or how well he knows certain specific words. It is especially important in studying technical vocabulary in school subjects to know, not what percentage of a list of words the pupil knows, but which individual words are likely to cause him difficulty. Information of this sort must be dependable, as it is the basis for remedial instruction. Several types of vocabulary test have been developed for diagnostic purposes, and some research has been reported which demonstrates advantages and disadvantages for those techniques. It is the purpose of this paper to discuss certain problems of diagnostic vocabulary testing and to evaluate the techniques available for this purpose. Conclusions are based principally on previous investigations and a logical analysis of the test forms themselves.

Types of word knowledge. Perhaps the leading source of confusion in this field is lack of agreement as to what "knowledge of a word" consists in.[2] Knowledge may be taken to imply ability to understand the word in recall or recognition. The student may encounter the word while reading or listening, and, in order to understand the discussion, must recall from memory the experience to which the term refers. Alternatively, he may encounter an object or event (such as the expression $x^2 - 4 = y$) and, in reacting, would find it useful to name the event with the proper concept (here, *quadratic equation*) so that he can use in his reaction whatever knowledge he has about that concept.

Various sorts of behavior that may be called for in understanding a word may be summarized in terms of these questions:

(1) Can the pupil define the word? This may, and is often understood to, indicate mastery of the term as a *generalization*.

(2) Can he recognize that an illustration of the word, as commonly employed, is properly named by that word? That is, can he select, in ordinary situations, those to which the word is appropriate? We may refer to this level of behavior as *application*.

(2) Can he recall the different meanings the word may have in different contexts? Further, can he distinguish which of these meanings is appropriate in a given context? For example, does the word *square* convey different meanings in the phrases "a square table", "seven square feet", "the square of seven", and so on? This aspect of word knowledge, which has been emphasized in writings on semantic problems of understanding, which we shall call *breadth* of meaning.

(4) Can he apply the term correctly to all possible situations—even unfamiliar ones—where its standard meaning would apply, and recognize its inapplicability to situations where it does not apply? For example, does his concept of *nation* recognize only that the United States is a *nation*, that New York and Latin America are not? Or can he further make such finer decisions as determining whether Canada, the Holy Roman Empire, Iceland, or the Confederate States would be so classified? This aspect, *precision* of meaning, is perhaps the most crucial for diagnostic testing, especially when vocabulary is used in a technical sense.

(5) Does the student, having the ability to use the term, actually make use of the concept in his thinking and discourse? This aspect we may call *availability*.

Many of these aspects have been distinguished by previous writers. Thus Dolch has commented:

> We need first to point out two extremes: a word as a symbol that can be put to *some* use in thinking or communication, and the word as representing the total fullness of meaning which human experience may associate with it. Between the two extremes there are naturally all degrees of meaning for different individuals. These degrees of "richness of meaning" are all extensions of the specific use to which the symbol is put when it first becomes a real word at all in thinking or communication.[4]

Later writers have followed the lead of Dolch in referring to "richness of meaning," which does not distinguish between breadth and precision as defined above. Since it is necessary to consider these separately in devising tests, the terminology suggested here appears advantageous. It should be noted that listing these aspects separately in this discussion does not imply that they are independent; in classroom work several aspects may appear simultaneously. It is not of value to discuss which type of knowledge is most significant for teaching; the value of a distinction lies in the fact that a clear

understanding of what behaviors are involved makes it possible to develop instruments to test those behaviors, and, conversely, to eliminate confusion by discussing just which of them any test measures.

Relation of types of word knowledge. Workers in this area have debated whether the testing of various aspects of word knowledge is possible or desirable; others have avoided testing for breadth and precision because of interest in estimating the pupil's total vocabulary rather than in diagnostic analysis. Thus Dolch[4] has stated that testing "richness of meaning" is impracticable, and he and others have confined their tests to words having single meanings. Such a procedure of course limits the diagnostic value of the test. Seashore and Eckerson, also interested in a survey test, comment, "Among professional linguists it might be worth while to insist upon very fine discriminations in alternative meanings [in the multiple-choice form], but for our purposes in a general test it is sufficient to discriminate the commonest meaning from those which are definitely incorrect but which might be sources of confusion to non-specialists in the field."[14] While this is appropriate to their purposes, it will readily be seen that for the teacher working with technical vocabulary, it becomes very important to test the student's ability to distinguish a correct from an almost correct meaning, in order to know the range of situations in which he can use the term without error.

Annen[14] found that several forms of vocabulary test (checking words known, defining, giving in a sentence, multiple-choice) gave similar results when the average vocabulary of a group was estimated by each. These data, according to Seashore and Eckerson, "suggest that knowledge of a word is ordinarily 'all or none', that is, if you can recognize it at all you can use it." This comment is likely to be misleading if it is taken to mean that the person who knows the word well enough to recognize it at all can also use or understand it in any situation. Seashore and Eckerson recognize that Annen's data are inconclusive. In comparing the four tests with each other, the error of comparing group averages instead of individual performances on individual words appears to have been made; obviously, if we are to know whether the student who recognizes any given word can also use it, the latter type of study is needed. Since the correlations between scores of individuals on the various tests were only about 0.40, the deduction made by Seashore and Eckerson appears questionable.

One might be tempted to draw similar conclusions from an experimental study by Lovell,[11] in which he reported that "richness of vocabulary (multiple meanings) is fairly closely related to . . . knowledge of single commonest meanings, so that for general purposes we may estimate one from the other." While this would seem, on the surface, to indicate that it is unnecessary to consider "breadth of meaning" in vocabulary research, an analysis of Lovell's procedure does not support such an inference. A multiple-choice test based on a representative sample of a dictionary was used to determine whether the student could understand the commonest meaning of each word included. To determine whether the student understood other possible meanings for those words, he was asked to use each in a sentence "which would illustrate the desired meaning as suggested in a cue meaning." Lovell suggests that a multiple-choice test might also serve this purpose if the labor of preparing it were justified. The total scores of each student on the commonest meanings test and on the multiple meanings test had a correlation of 0.68. This shows that the total number of words known by a student, in their commonest meaning, is correlated with the total number of meanings known for a list of words. This is not surprising, since when a word is used in two different ways, one merely has two different words spelled the same way. To estimate a student's total vocabulary, it is conceivably immaterial whether the words used as a sample for testing are unique in spelling or include homonyms. It is essential to note one important point which Lovell's treatment of breadth of meaning does not explore. The data obtained indicate only that the total number of commonest meanings is related to the total number of multiple meanings known, and give no diagnostic comparison. One cannot, then, conclude that for any single word, the extent to which a student understands all meanings of that word may be estimated from his knowledge of its commonest meaning.

On the other hand, there is considerable reason to believe that the aspects of word knowledge described above should be separately considered. Dewey illustrates how words may take on more or less precise meanings:

> Popular concepts, like the ordinary one of a dog, are based on fairly obvious qualities, qualities that anyone having the normal use of his senses can readily perceive. But these popular concepts do not take us very far; they can be extended or generalized to include cases outwardly different only at great risk. They are responsible for such generalizations as calling a bat a "bird", and a whale a "fish". They not only lead us astray, but they also stop

far short of such generalizations as those that are basic and almost commonplace in science . . . And it is this latter kind of concept that furthers discovery, invention, and the control of the forces of nature.[8]

Eskridge,[6] in his study of the growth of knowledge of geographical terms, provides clear evidence that concepts may be known precisely, hazily, or not at all. He distinguishes between basic (exact) meanings and associated meanings, and demonstrates that the latter disappear as the student learns the words more thoroughly. Eskridge obtained data by administering tests of several varieties, including short essay forms. Students able to respond successfully on one form were frequently unable to obtain credit for the word on another test. Eskridge notes that this should not be attributed to lack of validity in the tests, but rather to the fact that they measure different meanings or different degrees of meaning. "According to this view," he says, "the three different methods of testing employed really measured three different ways of knowing each term." This may imply, for example, that a student has only one meaning for the word, a meaning which leads to successful behavior on only some tests. A child with the concept of *bird* as "that which flies" might give a passable definition, might successfully give examples of the term, and yet might fail in a test which demands discrimination between true birds and insects or airplanes.

In view of Annen's finding that asking the pupil to check the words he understood gave results similar to those from more complex tests, the author has made a simple experimental study of the diagnostic value of the checklist with technical vocabulary; other similar studies making a comparison of diagnostic results from other test forms might be worthwhile. Fifty-nine terms were chosen at random from the technical vocabulary of general psychology, many of them words also found in physiology or other subjects. One nonsense word, *denomia,* was added, and the list arranged alphabetically. On the first day of the semester, two classes in general psychology were asked to "place a check in front of every word whose meaning you understand, as it would be used in psychology." Two students checked *denomia.* The next day, without previous warning, seven words from the list were presented to the groups with instructions to tell in twenty words, or thereabouts, what the word means, attempting every item. Students were asked to indicate if any word had been looked up between tests. It was understood that neither test had any bearing on the student's grade in psychology. The seven words for the second test (different in each group) were selected

to represent positions evenly spaced on the scale from those most frequently checked to those least checked. Papers for the essay test were scored on a five-point scale from zero to four, using semi-objective criteria. The bi-serial correlation for each word was computed and converted to its product-moment equivalent. The correlation between the student's score on the word in the essay test and whether he had checked it the first day ranged from 0.74 *(cutaneous)* to 0.05 *(reflex)*. The only other correlations greater than 0.30 were 0.50 *(blind spot)*, 0.43 *(adrenin* or *adrenalin)*, 0.38 *(telepathy)* and 0.35 *(aptitude)*. One group contained forty cases; the other, sixty-five. Since even these correlations are low, it must be concluded that marking a word on a checklist does not indicate understanding of the word. Correlations were especially low on terms like *maturation,* which was confused with *maturity,* and *probable error,* which is composed of non-technical words and gave rise to such guesses as "something that is probably wrong." As in Eskridge's study, many students passed words on the essay test which they had left unchecked on the first list.

Dolch[5] has studied the effectiveness of the checklist as a device to determine the relative difficulty of words (which is a major objective of diagnostic testing). He concluded that the checklist is not satisfactory, used by itself. Tilley,[16] however, compared the multiple-choice and checklist methods, using non-technical vocabulary, and concluded that the latter is sufficiently valid for determining relative word difficulty. The present data confirm Dolch's finding that the checklist is not dependable for this purpose, at least with these words. The frequency with which a word was checked was related only roughly to the mean score of the group on that word in the essay test. For *environment,* the percentage checking was 100, the mean score, 3.00; these values for other words were: *adolescence,* 100, 2.69; *reflex,* 91, 2.34; aptitude, 90, 2.18; *mental age,* 77, 2.00; *inhibition,* 72, 0.92; *adrenin,* 70, 2.02; *telepathy* 62, 2.62; *perspective,* 60, 0.38; *maturation,* 54, 1.03; *blind spot,* 47, 1.35; *probable error,* 37, 0.22; *socio-economic rating,* 20, 0.98; and *cutaneous,* 15, 0.57. Data from both groups have been mixed here; this appears to introduce no error.

Methods of testing word knowledge. Having demonstrated that there is reason to believe that various types of test measure various aspects of word knowledge, it then becomes important to see which testing procedure is appropriate for each level. It is relatively simple to test the first aspect, generalization, by asking the child to define the term, but this measure may

be invalid if the child has a correct concept yet is unable to express that concept in words. Another approach is the objective form, usually multiple-choice or matching, where the child recognizes one of several definitions as correct. The great weakness of testing at this level for diagnostic purposes is that the behavior shown in non-functional. The child may have learned a definition without having an understanding of it; he may then lack ability to use the word.

Much of previous vocabulary testing has been done at the application level. The form most used is multiple-choice, in which the student is to select which of several choices best illustrates the word given. The essay form, calling for illustrations of the concept, may also be used. The free recall method in which the student is asked to identify a picture or object with the correct word is more appropriate for young children with a small vocabulary than for those who can command many terms which could legitimately apply to the illustration. An objective form of this procedure has been successfully used at an advanced level by Faulkner and by Stone and Webb. To test students' knowledge of such technical terms in art as *realistic, abstract, dramatic, joyous,* and *static,* Faulkner[7] presented pictures or art objects and asked the students to indicate which of these words were illustrated by each. Stone and Webb[18] presented a list of terms in the field of conservation of pupils, together with a picture showing a conservation practice, and required the student to indicate which of the terms were illustrated in the photograph. About twenty-four words were presented in the list for each picture, twelve being keyed as correct.

Breadth of meaning has been given increasing attention as a result of recent interest in the connotations and shades of meaning of words. Woody[20] describes two devices for testing the number of meanings the student attaches to a word. Lovell's work also was directed toward determining how many meanings for a given word the student can recall or recognize. Tyler[17] has suggested the master list device for testing the ways in which the student understands technical words. Several words to be tested are placed in a numbered key list. Several definitions or descriptions illustrating various meanings of each word are placed in random, mixed order. The student marks each definition with the appropriate number. Tyler illustrates the use of this procedure in geography and botany; a later test by Koopman[10] has applied the technique to a large number of technical terms in the social

studies. These tests appear to have the weakness that, although determining how broad is the student's meaning for the word, they fail to determine whether he selects the appropriate meaning for any given context; a somewhat similar criticism has been voiced at length by Zahner.[21]

Precision of meaning—the extent to which a student's idea of a word is precise rather than vague—has been studied little in past tests. One explicit attempt to measure the exactness of concepts is reported by Atwell and Wells.[1] Their device requires that the student, given several definitions which are nearly correct, and one which is perfect, must make a fine discrimination between the choices; this is in contrast to the usual test where only one definition is at all correct. A rather similar test is used by Schaaf[18] to measure the pupil's understanding of algebraic terms by requiring him to choose, among several more or less close synonyms, that which most nearly represents the same meaning. The principal weakness in both these tests, as an attempt to measure precision of meaning, is that knowledge of the word is credited as right or wrong, on an all-or-none basis. Whereas these tests merely determine, with very difficult items, whether the student thoroughly understands the word or has not yet reached such a high level, it would be more valuable to know just how thorough or how incomplete is his concept. This would require a device on which a student can obtain credit along a scale from complete ignorance to complete understanding.

No procedure for testing knowledge of words from the standpoint of availability is known by the writer save that of analyzing a sample of the pupil's speech or writing to note the words used. This is illustrated in a study by Phipps,[12] who had children write answers to questions on history, and tabulated the words used by each to express himself.

Relative merits of test forms. Having thus summarized the sorts of test used at each level, we may look at the advantages and weaknesses of the various forms of test when used for diagnostic purposes. The forms most used are the recall types (written or oral), multiple-choice or matching to select either a correct definition, a synonym, or an example, and the check-list. The multiple-choice or completion type in which the word fitting a given definition is sought may also be employed. It will be noted that in most of these only one response is made by the student for any word, so that the information obtained about his knowledge of that word is likely to be insufficiently reliable for inventory purposes. Other criticisms of the

types of test now in use have been made by Dolch,[4] Seashore and Eckerson,[11] and Sims.[15] It is unnecessary to review these criticisms at length; common faults noted are the possibility of obtaining credit by guessing or by eliminating wrong responses, the stressing of verbalization instead of understanding,[8] and allowing full credit for partial or very hazy knowledge of the word. Weeks[19] and Atwell and Wells[1] demonstrated that in the multiple-choice form the student's choice is greatly affected by the position of the correct response; this of course hinders diagnostic use of the test.

Dolch,[4] on the basis of a logical analysis, recommended use of the matching technique. Sims studied the value of tests for survey, not diagnostic purposes. He concluded that the checklist technique is unsatisfactory and that on the basis of intercorrelations the multiple-choice, matching, and interview methods are valid, the two last-named being superior. Kelley compared five devices (multiple-choice of synonyms, recall of synonym or near-synonym, same-opposite-neither, matching, and multiple-choice with the stimulus word used in a sentence), and concluded that the matching and the first of the multiple-choice methods are the best, but that validity coefficients are too low to justify dependence on any of them. Validity was determined by correlating each test with a test in which the pupil was asked to use the word in a sentence. Eskridge's interpretation of similar data, quoted above, casts some doubt on this conclusion.

The difficulty of verbalism, or parroting words without understanding, appears to be inherent in any procedure requiring knowledge of definitions, whether a matching test, a master list test such as Tyler proposes, or the test of Atwell and Wells. It is possible to minimize the difficulty by phrasing definitions in a different manner than that of textbooks. It is possible, however, that the student who knows the word may fail to pass the item because of lack of understanding of words or phrasing used in the definition. The most promising development to avoid this problem, the work of Faulkner and Stone and Webb with pictures, will probably be difficult to adapt to many abstract words.

Another difficulty which is inherent in almost any test requiring that the student make only one response for each word is that such responses do not necessarily indicate whether he knows the word. An example will show that even a simple situation of the multiple-choice type where one is certain

that there has been no guessing offers possibility of error in diagnosis. Consider the exercise:

spade *shovel* *trowel* *house* *broom* *tool*

The child who choose *trowel* as the best synonym for *spade* does not thereby show that he misunderstands *spade.* Perhaps he has a false meaning for *trowel;* perhaps he has a false idea of *shovel,* and, rejecting it as an answer, is forced to take a wrong choice; or, conceivably, his concept of *spade* is so precise that he finds *shovel* a poor substitute and selects *trowel,* with which he is unfamiliar, by elimination. In any case, he fails to receive credit despite precise knowledge of *spade.* The converse, where credit is given despite lack of knowledge, can arise at least as frequently. Similar errors can also rise in the matching test, especially if the pupil is instructed to use each choice in the second column only once. If the student makes one error, he is almost certain to make another error in consequence of it. The interlocking of responses on such a test makes on uncertain whether the student's errors are due to ignorance on every item missed, are due to misunderstanding of one or more of the definitions offered for matching, or are due to lack of knowledge of just one of the words. Kelley found that the average score on a matching test was approximately twenty per cent greater than on a multiple-choice test, when the same words were used and the correct choices in both tests were the same; this suggests that solving by elimination is prominent. For a valid diagnostic test, it is essential that the exercises be independent.

One form of test which does succeed in minimizing all these difficulties, but which requires considerable care if data are to be treated quantitatively, is the interview. In this method it is possible to ask the student several questions on each word, if necessary, to make certain that his success represents knowledge of the word rather than some other factor. The tests described using pictures also appear worth further study; while they appear valuable at the application level, they are not yet adapted to testing for breadth or precision of meaning.

With these exceptions, it appears that diagnostic vocabulary testing requires more valid instruments than are now available. It appears especially important, if diagnosis is to be accurate, that the possibility of obtaining credit without knowing the word be reduced to a minimum. Probably the most successful way to attain this is to ask several questions about each word;

the student who can respond successfully to all of them is far more likely to know the word than whoever succeeds on his one attempt. It is hoped that a recognition of the weaknesses of present tests, and of the types of information about word knowledge they do not provide, will lead to development of superior tests, especially to study breadth and precision of meaning. A subsequent article will suggest a procedure which may be useful in improving measurement of precision of meaning.

REFERENCES

1. Atwell, C. R., and Wells, F. L. "Wide range multiple choice vocabulary tests," *Journal of Applied Psychology,* XXI (October, 1937) pp. 550–555.
2. Dale, Edgar. "Difficulties in vocabulary research," Educational Research Bulletin, X (March 4, 1931) pp. 119–122.
3. Dewey, John. *How We Think.* (New York: D. C. Heath and Co., 1933.) pp. 180–1.
4. Dolch, E. C. *Reading and Word Meanings.* (New York: Ginn and Co., 1927), 129 p.
5. Dolch, E. C. "Testing word difficulty," *Journal of Educational Research,* XXVI (September, 1932) pp. 22–27.
6. Eskridge, T. J., Jr. *Growth in Understanding of Geographic Terms in Grades IV to VII.* Duke University Research Studies in Education, No. 4, (Durham, N. C.: Duke University Press, 1939) 67 p.
7. Faulkner, Ray. "Evaluation in a general art course." *Journal of Educational Psychology,* XXXI (October, 1940) pp. 481–506.
8. Judd, C. H. *Educational Psychology.* (Cambridge: Houghton–Mifflin, 1939) pp. 163 ff.
9. Kelley, V. H. *An Experimental Study of Certain Techniques for Testing Word Meaning.* University of Iowa Studies in Education, IX, No. 1, (Iowa City: University of Iowa, 1934) pp. 53–94. Also, *Journal of Educational Research,* XXVII (December, 1933) pp. 277–282.
10. Koopman, M. O. Test, *Social Science Concepts,* Form A. (Chicago: Progressive Education Association) 8 p.
11. Lovell, George D. "Interrelations of Vocabulary skills: commonest *versus* multiple meanings," *Journal of Educational Psychology,* XXXII (January, 1941) pp. 67–72.
12. Phipps, William R. *An Experimental Study in Developing History Reading Ability with Sixth Grade Children through Development of History Vocabulary.* (Baltimore: John Hopkins Press, 1940) 73 p.
13. Schaaf, W. L. "Testing the clarity of mathematical concepts," *School Science and Mathematics,* LXI (October, 1939) pp. 651–656.

14. Seashore, R. H., and Eckerson, L. D. "The measurement of individual differences in general English vocabularies," *Journal of Educational Psychology,* XXXI (January, 1940) pp. 14–38.

15. Sims, V. M. "The reliability and validity of four types of vocabulary tests," *Journal of Educational Research,* XX (September, 1929) pp. 91–96.

16. Tilley, H. C. "A technique for measuring the relative difficulty of word meanings among elementary school children," *Journal of Experimental Education,* V (September, 1936) pp. 61–64.

17. Tyler, R. W. *Constructing Achievement Tests,* pp. 41–52, "The master list as a device." (Columbus: Ohio State University, 1934.)

18. Webb, E. M. *Natural Resources in the Public Schools.* Unpublished master's thesis, State College of Washington, 1941.

19. Weeks, A. L. "A vocabulary information test," *Archives of Psychology,* XV, No. 97, 1928, 71 p.

20. Woody, Clifford. "Attempts at measurement of meaningful experiences as a factor conditioning achievement in reading," *Peabody Journal of Education,* XVI (November, 1938) pp. 180–191.

21. Zahner, L. C. "The testing of comprehension," *Educational Record,* Supplement No. 13, XXI (January, 1940) pp. 71–89.

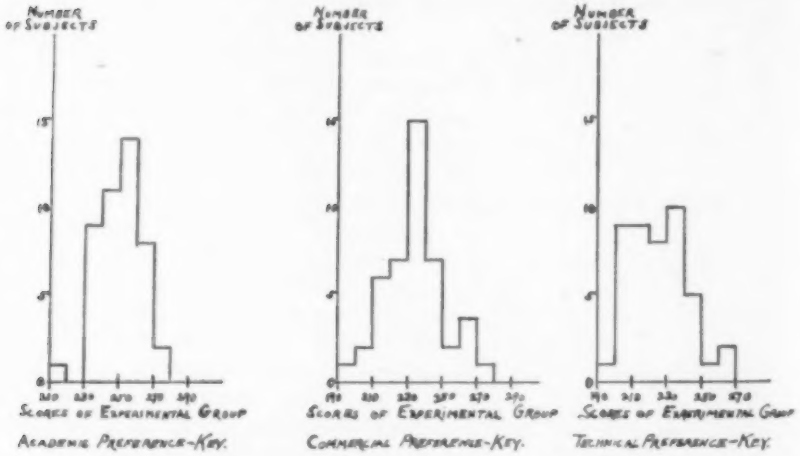
A GRAPHICAL DEVICE FOR COMPARING THE FORM OF A GIVEN DISTRIBUTION OF TEST SCORES WITH THE FORM OF THE STANDARD DISTRIBUTION FOR THE TEST

ALEXANDER W. MORRISON
Columbia University

Editor's note: The author presents a graphical device for comparing the distribution of test scores for a given group with that of a standard group.

IN WRITING-UP a study in which the Garretson-Symonds High School Interest Questionnaire had been administered to forty-five adolescent male subjects, it was desired to show graphically the relationships of their scores on the three preference-keys of the questionnaire to the distributions of the scores on these same keys made by the standardizing groups. The customary frequency distributions of the scores, shown in Chart I, could not show these

CHART I

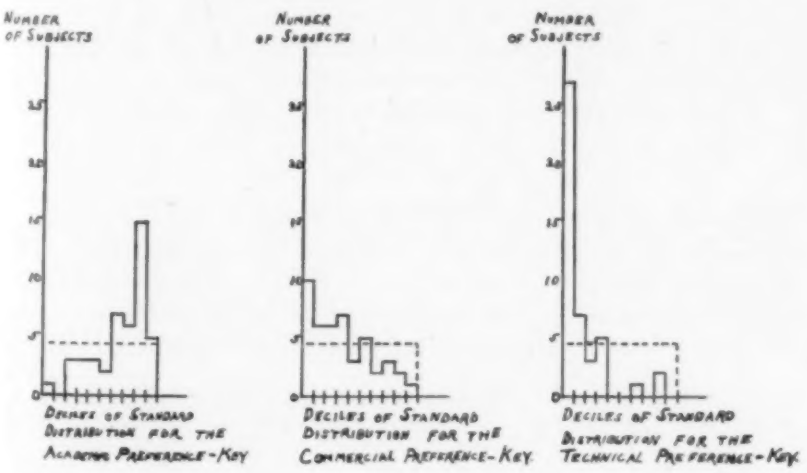


relationships without the means and the standard deviations, or suitable substitute measures, of the various sets of scores for the experimental group being equated to the corresponding measures for the scores of the standardizing populations, since these measures of the two populations for each preference-key would otherwise be entirely unrelated to one another.

However, the score values which defined the limits of the various decile-classes of the standardizing distributions were available and it appeared that it would be justifiable to assign each preference-score of a member of the experimental group to the decile-class of the corresponding standard distribution into which it might fall, since the maximum and minimum score values obtainable on the three preference-keys were known and limited the possible variations of the scores of both the standardizing and the experimental groups. For instance, the maximum score obtainable on the Academic Preference-key of the questionnaire is 327, and the minimum score is 159, while for the Commercial Preference-key the maximum and minimum scores are 366 and 124, respectively, and on the Technical Preference-key they are 316 and 137. Thus, the possibility of score variation differed for the three methods of scoring, but for both the standardizing and the experimental group scores derived from a given preference-key the possible variation was the same.

After the scores had been classified according to the decile-classes of their respective standard distributions it was clear that the tendency of the scores of the experimental group was to be concentrated in the upper or the lower decile-classes of the distributions obtained from the standardizing groups, as can be seen in Chart II. That is, more than fifty per cent of the

CHART II


NUMBER OF SUBJECTS
DECILES OF STANDARD DISTRIBUTION FOR THE ACADEMIC PREFERENCE-KEY
NUMBER OF SUBJECTS
DECILES OF STANDARD DISTRIBUTION FOR THE COMMERCIAL PREFERENCE-KEY
NUMBER OF SUBJECTS
DECILES OF STANDARD DISTRIBUTION FOR THE TECHNICAL PREFERENCE-KEY

experimental group had academic preference-scores *above* the median of the selected standardizing population while, on the other hand, this same group of subjects had more than fifty per cent of both their commercial preference-scores and their technical preference-scores *below* the medians of the standardizing groups for these preference-keys. In fact, the marked dearth of interest in technical subject-matter among these boys is probably brought out more strikingly for the non-technical reader of a report of this study by means of Chart II than could be done by any other method.

If the distributions of the scores of the experimental group on the various preference-keys had duplicated in graphical form those of the standardizing groups, and the corresponding means and standard deviations had been equal, the form assumed by the charts of the experimental group scores in terms of decile-classes would have been that of a perfect rectangle, as shown by the dashed lines on the figures in Chart II.

A STUDY OF THE RELATION OF RETARDATION IN READING TO TEST PERFORMANCE ON THE REVISED STANFORD–BINET, (FORM L)*

SAMUEL KAVRUCK
The New York State Training School for Boys
Warwick, New York

Editor's note: The author finds that the relation between reading ability and performance on part of the Stanford–Binet (Form L) differed with the vocabulary and abstracts of the Stanford–Binet giving the highest correlations.

PURPOSE OF INVESTIGATION

THE purpose of the investigation was to determine among delinquent boys the relationship of performance on the Stanford revision of the Binet Test and achievement in reading. The study involves 100 boys in the New York State Training School for delinquents. The boys were divided into two groups, fifty in each group, and matched according to chronological age and I.Q. as determined on the revised Stanford–Binet Test, Form L. One group, however, attained the expected standard of achievement in reading; the other group attained a standard which was at least two years below the expected standard, according to mental age and performance on the Stanford Achievement reading tests. The chronological ages of the boys ranged from 13–0 to 16–0 years, with I.Q.s from 66 to 112. The median I.Q. for the retarded group was 90.1 and for the normal group, 90.0. There were 25 Negro and 25 white boys in the retarded group, and 24 Negro and 26 white boys in the normal group. The problem, then, was to determine whether or not significant differences were shown in the responses of the two groups on the various parts of the revised Stanford–Binet, Form L.

PROCEDURE

Successes at each year level of the Binet were recorded for each subject. Both groups scattered from Year VII to Superior Adult III. Standard errors of the percentages for each item were calculated and critical ratios established. A critical ratio of 3.0 or above was used as the criterion of a significant difference. Significant differences were established only in year levels

* Studies From The New York State Training School For Boys, Warwick, N. Y. I wish to thank Dr. Richard L. Jenkins and Mr. Daniel P. Clarke of the staff for helpful criticism.

TABLE I

DIFFERENCES BETWEEN THE NORMAL AND RETARDED READERS ON THE REVISED STANFORD–BINET, FORM L

Year level and subtest on the Rev. Stan.-Bin. (L)	Items in which the Normal Readers excelled	Critical* Ratio
XII–1	Vocabulary	4. 5
XII–6	Minkus Completion	3. 2
XIII–5	Dissected Sentences	5. 0
XIV–1	Vocabulary	5. 0
XIV–6	Abstract Words	3. 8
	Items in which the Retarded Readers excelled	
XI–1	Memory for Designs	3. 9
XI–4	Sentence Memory	3. 0
XIII–6	Bead Chain from Memory	4. 0
	Items showing little or no difference	
XIV–5	Orientation: Direction I	Below 3. 0
Av. Adult–2	Codes	Below 3. 0
Av. Adult–4	Arithmetic Reasoning	Below 3. 0

* Critical Ratio: is the difference between two comparable statistics divided by the standard error of that difference.

XI–XIV inclusive, since there were too few failures in levels VII–X and too many failures above year level XIV. The mean number of levels of scatter for the retarded group was 6.9 and for the normal group, 6.8.

RESULTS

Significant differences in the responses of the two groups on the various parts of the revised Stanford–Binet, Form L, are shown in Table I. The retarded readers were inferior to the normal readers in vocabulary, in defining abstract words, in the Minkus completion and in assembling dissected sentences. They excelled in memory for designs, in sentence memory and in constructing a bead chain from memory. The retarded group would be expected to do less well than the normal group on Stanford–Binet items which involve reading ability. Since the two groups were matched for C.A. and I.Q., it is obvious that the poor readers, by virtue of being matched for intelligence test score, would have to excell in test items not dependent on reading ability. It is of great interest, however, that the only significant differences in favor of the retarded readers were on test items involving memory. Whether this is a compensatory mechanism on the part of the

poorer readers, is an interesting speculation. That the blind compensate for their loss of sight by the development of skill in the use of touch and hearing is well known.

The retarded group showed no difference in the Arithmetic Reasoning item even though it requires that the subject read it. Here the vocabulary is simple enough for the poor readers. Although one would expect the retarded group to scatter over more levels on the Binet than the normal group, this was not the case.

A final point is of interest. Of the items on which the normal group is superior, the vocabulary and abstract words do not require that the boy read them. Instead, they seem to be associated with good reading ability. It is expected that a boy with reading ability would tend to build up his vocabulary, whereas this development would lag in the boy with a reading disability.

Research Abstracts and Bibliographies

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

WRIGHTSTONE, J. WAYNE. *Broadcasts versus Transcriptions in the Classroom.* Bulletin No. 49. Columbus, Ohio: Evaluation of School Broadcasts, Ohio State University, 1941. Pp. 11.

In a study conducted with seventh and eighth grade pupils in Bronxville, New York, no significant difference was found between the educational effectiveness of direct radio broadcasts and transcriptions of the same over a public address system.

ZERAN, FRANKLIN R. *Manual for Occupational Studies Leaflet.* Occupational Information and Guidance Service, Misc. 2922. Washington, D. C.: Office of Education, Vocational Division, 1941. Pp. 16.

A manual to assist the counselor in the most efficient use of the *Occupational Studies Leaflet.*

EDUCATIONAL ADMINISTRATION AND SUPERVISION

FERGUSON, HAROLD, and HOVDE, HERMAN O. "Improving Teaching Personality by Pupil Rating," *School Review,* L (June, 1942), 439–443.

The report of a study in which students and fellow teachers rated an instructor on twelve personality traits. The advantages and short-comings of such a procedure for improving teaching personality are analyzed.

SUPERVISION, CURRICULUM, AND INSTRUCTIONAL PROCEDURES

GRAY, WILLIAM S. "Education of the Gifted Child: With Special Reference to Reading," *Elementary School Journal,* XLII (June, 1942), 736–744.

The author discusses the characteristics of bright pupils and proposed adjustments for them. Emphasis is placed on adjustments in reading programs.

HUNTER, GEORGE W., and SPORE, LEROY. "Science Sequence and Enrollments in Secondary Schools of the United States," *Science Education,* XXVI (February, 1942), 66–77.

Second half of the report of a survey of science enrollments in schools in the United States and several Canadian provinces accompanied by tables and figures showing the distribution for the various science divisions by grades and by geographical areas.

JOSEPH, ALEXANDER. "Developing a Sourcebook of Extra-Curricular Activities in Physical Science for Senior High Schools," *Science Education,* XXVI (February, 1942), 84–93.

Six criteria for extra-curricular activities in the physical sciences selected by science educators and science club leaders were used in evaluating 2358 activities in physical science. One hundred and eighty-

five experiments, experiences, and projects were selected for extra-curricular activities.

KARPINOS, BERNARD D., and SOMMERS, HERBERT J. "Educational Attainment of Urban Youth in Various Income Classes. I," *Elementary School Review,* XLII (May, 1942), 677–687.

Part I of a report on a survey of 83 cities in which educational status is compared directly to annual family income. Educational attainment of both white and colored youth is analyzed and compared

LUCKIESH, MATTHEW, and MOSS, FRANK K. "The Task of Reading," *Elementary School Journal,* XLII (March, 1942), 510–514.

A study of eye fixations during reading and the relative changes occurring in the processes of reading during an hour of reading. Implications for school activities are suggested.

RICHTER, CARLTON E., and PARR, F. W. "Remedial-Reading Instruction in Oregon Secondary Schools," *School Review,* L (May, 1942), 368–380.

A survey of remedial reading practices in Oregon public schools reveals the growing interest in the problem, methods used to diagnose and treat reading deficiencies, and the need for a more extensive remedial-reading program.

STEVENS, LEILA. "Poetry Presentation of Second- and Third-Grade Readers," *Elementary School Journal,* XLII (March, 1942), 534–539.

An analysis of twenty-two second and third grade readers reveals the fact that approximately four per cent of the reading material is poetry, and that there is relatively more poetry in the second grade books. The author also analyzes the subject matter and the authors of the poems.

WILLEY, ROY DEVERL. "Arithmetical Processes Needed by Children," *Elementary School Journal,* XLII (March, 1942), 524–527.

Selected classroom teachers in urban and rural schools of Santa Clara County, California, recorded 2484 arithmetical problems encountered by children enrolled in kindergarten through the sixth grade. The arithmetical processes required to solve these problems are compared with those taught.

MEASUREMENT, STATISTICS, AND SCIENTIFIC TECHNIQUES

GOOD, CARTER V. "Sources of Problems for Research," *Elementary School Journal,* XLII (June, 1942), 745–754.

The author discusses seven sources of problems for research: specialization, instructional program pursued, program of reading, analysis of an area of knowledge, consideration of existing practices and needs, repetition or extension of investigation, and "offshoots" of studies under way.

REID, SEERLEY. "Respondents and Non-Respondents to Mail Questionnaires," *Educational Research Bulletin,* XXI (April, 1942), 87–96.

The author finds that first returns from questionnaires may represent a biased sample which fact is not revealed until results from follow-up questionnaires are analyzed.

Research News and Communications

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Long-Range Planning for Education.—The Office of Education, in accordance with responsibilities imposed by the Act of Congress establishing it, is normally engaged in research studies and investigations, and the preparation and dissemination of reports on the same, that have for their purpose assistance to state and local communities in planning for the "establishment and maintenance of efficient school systems." In view of present exigencies in our social, economic, and civic life that now affect—and will affect for an indeterminate period—provisions for education, the Office is exerting a special effort to render assistance to schools in planning programs to cope with changed and changing conditions. To this end, the Office is undertaking the analysis of problems that will be basic for effective planning on the part of public school systems to meet the needs for adjustments during the present and post-emergency periods.

In order to have the counsel and cooperation of school officials in making studies on a national basis, the Office is working with the National Council of Chief State School Officers. The Executive Committee of that Council has met, at the request of the Commissioner, to consider the fields within which long-range planning for education is now seriously needed. Following that meeting a list of proposed subjects was sent to all state superintendents of public instruction with a request that each one indicate the problems of special importance to his state and the ones on which he wishes to cooperate. Preliminary steps have already been taken to make an initial survey of the school building needs throughout the country as a basis for developing long-range plans for schoolhouse construction.

The Commissioner of Education has appointed a committee of members of the staff to serve as the Office of Education Planning Committee and at the same time has invited a group of authorities in different fields of education to serve as a general advisory committee to the Office of Education Planning Committee.

Cumulative Records Study.—One of the frequent requests made of the Office of Education is: "Please send us a copy of a good cumulative record." Because of this continued demand, from time to time the Office has made studies of the forms of the cumulative records used by different school systems. Thus, it has been able to inform school systems desiring to inaugurate new cumulative record systems as to what others are doing. However, it was felt more and more as time passed, that school systems were perhaps too often adopting, without change, the record systems of others and that what was needed was a re-evaluation of the whole problem. To accomplish this, it was planned to do two things: first, to

make a more complete analysis of what schools are now doing by asking them, not only for their cumulative record forms, but also to what extent the forms have been used; second, and probably the most important step planned, is the appraisal of the validity of the items found on cumulative records used by the schools of the country and descriptions of how the cumulative records can be used by counselors, teachers, and administrators. The second step is being accomplished through the use of a committee composed of city research and guidance directors, state department supervisors, college professors, government sociologists, and health experts. It is hoped that, out of this work, the committee will be able to present to the country some recommendations which will enable school systems to develop functional cumulative record forms.

Not only may cumulative records be improved through this concerted action, but uniformity in reporting some of the information in cumulative records, information that all school systems collect in some form or other, may also result. This would be of especial value in the transfer of records from one school system to another. It is hoped that the easy transfer of records from one school system to another will be an incidental (and important) result of this study. To explore this possibility, the Office has issued a publication, based on the data of this study, called a *List of Cities and Counties Using Cumulative Records.* In this publication are listed cities and counties using cumulative records, together with the grades in which the records are used and the year the system was started.

Inter-American Friendship and Understanding.—In cooperation with the Division of Cultural Relations of the Department of State, the Office of Education, through its Division of Inter-American Educational Relations, administers a program of exchange fellowships and scholarships.

The Division of Inter-American Educational Relations also cooperates with the Office of the Coordinator of Inter-American Affairs in the distribution of educational publications of use to teachers and others interested in this subject. These publications include those issued by the Office of Education and by other agencies working in the field. This Joint Distribution Center, as it is called, is also prepared to answer questions with regard to the organization of units and courses of study, the preparation of programs, and the selection of useful books and other teaching aids.

Another project in process of development is the Inter-American Demonstration Centers. A number of schools throughout the country have been invited to work with the Office of Education in setting up an educational program emphasizing the better understanding and greater appreciation of the people of the other Americas.

In the Library Service Division of the Office, exhibits of materials of value to teachers have been prepared.

National Defense Publications.—The Office of Education has under way the publication of a series of twenty-five pamphlets on "Education and National Defense," eight of which were released during 1941, and six of which were in press as of January, 1942. Publication of the others has been in progress during the

current year. This series of pamphlets includes a group that gives special consideration to teaching and practicing democracy in various educational situations: "What Democracy Means" (for elementary schools), "Living Democracy in Secondary Schools," "Practicing Democracy in the College," and "Education under Dictatorships and in Democracies." Other pamphlets in the series deal with problems associated with democratic living in school and community.

During 1942 the Office of Education has attacked other problems made acute by the total war situation. The protection of children against war propaganda and needless strain is a responsibility of both parents and teachers. The most effective use of school time, with as little interruption as possible of the normal activities of the classroom is another challenge. The Office of Education is giving attention to the proper feeding of school children, precautions against possible air raids during school hours, effective use of school facilities in community organization, and evacuation procedures, if such become necessary for school children; and it expects to make available concrete suggestions that will help to safeguard the mental and physical health of both pupils and teachers.

In every project of this type, one encounters the problem of coordinating the services of all contributing agencies. At this critical time, when every one is eager to serve, organizations sometimes endeavor to meet needs quite outside their own fields of interest. The lack of coordination brings duplication, overlapping, and even conflict, while adequate planning means a stronger output and more effective results. It is hoped that during 1942 a better coordination of all available resources has been realized, looking toward better service on the part of each one in the interest of education for defense and for victory. The Office of Education is the leading agency for promoting this coordination of effort.

National Committee on Education and Defense.—The transition from a limited emergency to total war has challenged every institution of American life, education most of all. The material structures of life can be changed without permanent impairment. Social and spiritual values must be maintained for the future welfare of society, yet modified to render the greatest possible service in meeting the imperative needs of war.

It is this dual function of education, immediate and long range, that has challenged education, and it is the attempt to maintain perspective and judgment to meet this challenge that has been the major function of the National Committee on Education and Defense. Representing sixty national educational associations, it has direct entrance to every classroom and through the publications of its member organizations reaches every administrator and teacher on all levels. Through its sponsoring organizations, the American Council on Education and the National Education Association, there has been provided a channel for collective appraisal and coordinated action.

The National Committee on Education and Defense has operated through an executive committee and six standing committees, which are: Committee on Military Affairs, Committee on Teaching Materials, Committee on Vocational Training, Committee on Pre-Service Training, Committee on Women in College and

Defense, and the Committee on Inter-American Educational Activities.

The specific activities have included the calling of three national conferences on higher education, the preparation of statements of basic policy, and continual contact with the many agencies of government that directly affect education—the War and Navy Departments, Selective Service, the Office of Civilian Defense, the National Roster of Scientific and Specialized Personnel, the Civil Service Commission, and many others.

Through the work of Francis J. Brown of the American Council on Education and the other members of the Council staff, direct liaison has been maintained between these agencies and education, especially in the university and college field. The bulletin, *Higher Education and National Defense*, and the publication, *War Service Opportunities for College and University Students*, have supplied and, when desirable, interpreted information and developments emanating from Washington.

In the field of manpower, the Committee on Military Affairs has continually cooperated with National Headquarters of the Selective Service System, in making national surveys and in the preparation of memoranda regarding occupational deferment of students and faculty. In cooperation with the National Roster of Scientific and Specialized Personnel, the American Council on Education has conducted a national survey of available personnel, actual and potential shortages, and the training facilities of colleges and universities in the technical and professional fields. Through a conference of representatives of government and of higher education, channels were established for the cooperation of competing agencies for manpower. This is an illustration of but one of the many activities of the National Committee on Education and Defense and of the American Council on Education.

No one can predict the developments of the next months or years, but education has an established medium through which to seek cooperatively to face these developments realistically and to evolve policies that will make it possible for education to render its maximum service to the total war and to eventual peace.

This news item was forwarded by George F. Zook, president of the American Council on Education.

Test on the Effects of War.—Schools interested in developing pupil morale for meeting war hardships will be interested in a "Test on the Effects of War," designed to locate pupils whose morale is unsatisfactory and to identify war problems about which further instruction is needed. This test, prepared by Lee J. Cronbach, has been released by the School of Education of the State College of Washington, Pullman, Washington. It is planned for grades 10, 11, and 12, but may be used at higher levels. Seventy statements about conceivable future developments are presented, and the pupil is required to respond by indicating how likely he thinks each effect is. Responses are analyzed to determine how optimistic or pessimistic each pupil is. Since good morale depends on a realistic outlook and on planning for future developments, both the highly optimistic or complacent pupil and the highly pessimistic, panicky pupil are pointed out as cases for individual guidance. An item analysis of the responses of the group indicates those particular war problems about which pupils appear poorly informed.

The test is being made available as a professional service on a non-commercial basis to interested schools. For greatest value in planning the school program during wartime, the test should be given as early as possible. Question sheets, which may be used any number of times, sell for one cent apiece. Answer sheets, one of which is needed for every pupil tested, sell for five cents each. This charge covers the cost of producing the test and of a complete scoring service. All papers are scored, analyzed, and interpreted by the State Collge of Washington without additional charge.

The test has been standardized on nearly two thousand pupils in the state of Washington, tested during January and February, 1942. The reliability of the Optimism score, on which the principal interpretations are based, is .77.

Teacher Education.—The Commission on Teacher Education has endeavored to build its program on first things. It has furthered the cooperative process in every phase of its work and put its best energies into fostering a significant and mutually beneficial meeting of minds—as between staff members and administrators within a particular institution or school system, as between collegiate institutions and school systems as such, and among the several educational interests of whole states. It has likewise sought to emphasize throughout the supreme value of human personality—not only through the work of its experts on child development and teacher personnel, but also through its emphasis on group methods and self-evaluation at each stage of the unfolding enterprise.

The national cooperative study of teacher education reaches some fifty universities, colleges, and school systems in all major sections of the country. There are extensive programs on a statewide basis under way in three states, and assistance has been given to less comprehensive attacks on common problems in five or six more. The Commission has encouraged the institutions with which it is cooperating to share with others, as far as possible, the results and advantages of the contacts provided. When possible, it has welcomed to its collaboration center in Chicago and to its workshops representatives of colleges and school systems other than those with which it is most closely allied. As projects have matured and been reported, it has sought to make the facts known through mimeographed or printed pamphlets and in the pages of its monthly *Newsletter*.

In other words, the Commission has attempted to give a practical demonstration of democratic principles and to share its experience as widely as possible. While the war has brought confusing pressures and new problems of many kinds, there is no disposition now to change its main direction because of the national crisis. Members of the Commission and its staff have welcomed opportunities as they have been arising to assist the government in developing national policies, especially those affecting teachers and the education of teachers. They will welcome any further opportunities of this sort in the future. At the same time, their faith in the importance of what they originally set out to do in 1939 has not been impaired.

This statement was formulated by Helen E. Davis of the Commission staff.

Implications of the Stanford Social Education Investigation for Higher Education.—The work of the Stanford Social Education Investigation has pointed to the need for a greater consideration of the problems of experienced teachers now on the job. We cannot expect extended significant improvements in education to be made by beginning teachers or by the few teachers who now attend universities to secure higher degrees unless they are adequately aided and supported by the great majority of teachers in service. Hence it seems desirable that teachers colleges and colleges of education in universities give greater attention to in-service education, which involves the following problems and procedures:

1. A more effective use of interneship training as a regular part of the preservice education program.
2. A more careful follow-up of beginning teachers during their first experiences in the field.
3. The development of consultant relationships with school systems, so that members of college and university faculties can work directly with teachers on their problems in local situations.
4. Giving of encouragement and aid in the setting up of continuing programs of in-service education in local school systems, including local workshops.
5. Increasing the flexibility of the programs within colleges and universities, so that teachers can come in for short periods to do intensive work on their own problems, utilizing all the resources of the university both in personnel and materials.
6. Higher education could render a service to local school systems on a regional basis by circulating bibliographies, statements on experimentation, samples of newer materials such as recordings, kodachrome slides, motion pictures, and the like, and by supplying samples of organized teaching materials in the form of resource units in such rapidly developing fields as consumer education, vocational guidance, education for democracy, and the like.
7. The investigation has indicated the need for broader courses at the general college level, giving a background particularly in the social (including the arts), scientific, and psychological areas. There is also need for more self-dependence in the location of information on significant problems after students leave the university. Greater emphasis on the development of a variety of interests would be helpful.
8. There is need for a more careful study of the potentialities of students in college and a more adequate guidance program than now exists in most institutions of higher learning.

Many of these things are already being done by institutions of higher education throughout the nation. However, the experiences in the Stanford investigation indicate a much greater need for these types of service. Staff members of the study have experimented extensively with all of them and find that teachers are quite receptive to such aid and that the results are gratifying. Within the regularly organized programs of teacher preparation there is a need for the de-

velopment of a closer relationship between theory and practice. This should include greater cooperation between professional educators and teachers in supporting academic fields. It should also include more professional courses where educational psychologists, sociologists, and specialists in curriculum, guidance, supervision, and administration can work more closely together and direct the resources of their fields toward helping students to see how selected problems of education can be solved. This would involve a closer tie-up with the on-going life in the surrounding community and region, not only in the schools but with political, economic, cultural, and social leaders and groups.

The Stanford investigation has also stressed the need for more emphasis on the importance of successful social relationships in the classroom, and between the school and the community. Little attention has been given to this area in institutions of higher learning, but research is under way that should provide resources for individuals who are interested in the development of more effective relationships, both as a basis for success in teaching and as a major means for the fuller realization of the values of democracy.

This statement was prepared by I. James Quillen, co-director of the project.

Education of Science Teachers.—Educators are faced with a great responsibility as we approach the halfway mark of this century, which Vice-President Wallace has boldly asserted must become the "century of the common man." In order to enable all men to realize for themselves and their communities a rising scale of material and spiritual well-being, there must be development of latent resources throughout the world, increased productive power and exchange, and the maintenance of communities in a manner favorable to mental and physical health.

The agencies of ***general education*** (elementary, secondary, higher, and adult) must be enlisted to contribute a scientifically and politically literate citizenry capable of understanding and participating effectively in the planning that is essential to the carrying out of the war effort and to the building of a world-order of free and equivalent peoples. Such participation will involve two kinds of collaboration hitherto only imperfectly realized: collaboration with experts and collaboration with other peoples, both of which can begin at home and abroad simultaneously.

The Sub-committee on Teacher Education of the National Council of Science Teachers, working under the sponsorship of the National Education Association, has most appropriately, therefore, directed its attention to the pre-service and in-service education of teachers of science in general education. Since the work of scientific men has become of such central import to our changing culture, and since the objectives are to be validated continuously by democratically determined value-judgments, the integrated competencies of natural and social scientists are indicated as constituting the core of the general curriculum. Such an integration — again, Vice-President Wallace has suggested it— might well prove the distinctive contribution of democratic institutions to world-civilization. It would provide that both the scientific and the political procedures for solving problems are responsibly explored and fearlessly developed, "as men learn."

The general curriculum is proposed as good for all citizens, experts and non-experts alike. Professional scientists in training would, of course, continue into their appropriate specializations. For the teacher of science in general education, however, a new definition of possible specializations seems necessary rather than a continued pursuit of studies designed for medical, engineering, or research scientists.

Consequently, by way of illustration, the Sub-committee has proposed such areas as the following: the human life span; individual and group differences and similarities; the maintenance of personal and public health; the production, distribution, and consumption of materials and energy; the impact of technology on civilization; and the nature of the world.

More generally, the Sub-committee identifies such large educational (and social) objectives as the following: successful living in intimate relationships; successful collaboration between peoples and groups of people; successful philosophical orientation to the world and to society; successful participation in developing one's community (local, regional, national, world) as a good community in which to live and work; and successful participation in planning the full, stabilized employment of resources, techniques, and men the world around for achieving rising and equivalent material and cultural standards of living for all peoples. Educators will direct their efforts toward planning a curriculum within which science teachers and others may cooperate in work to achieve these and similar socially valid objectives.

Within this type of curriculum the science teacher, like other teachers, will be thought of as first of all a person; but he is also a professional worker with young people, striving increasingly to make "the community his laboratory, community problems his materials, scientific procedures his methods, and the whole of society . . . the area in which he works and lives."

Educational investigators may well be concerned with examining the means employed and with evaluation of successes that result from efforts to learn about the world as a civic unit and about the diverse groupings of people that inhabit it. There is indeed a challenge to institutions of higher education to use their resources to advance research and scholarship along the lines suggested.

This project has been described by S. R. Powers of Columbia University, Chairman of the Sub-committee on Teacher Education.

Junior College Terminal Education.—Recent extensive studies by the American Association of Junior Colleges, made on a nation-wide basis just prior to the dislocation of normal occupational and educational conditions resulting from the war, show that three-quarters of the young people who enter the more than six hundred junior colleges in the country do not continue their formal education beyond the junior college. Similar although less extensive studies in the senior colleges and universities have also shown that a large proportion of the entering freshmen in these institutions do not continue beyond the sophomore year.

It has been recognized for twenty-five years and longer that an important function of the junior college is to duplicate the work of the first two years of the university under more economical and often more favorable educational condi-

tions in smaller local units, for the sake of the student who will later transfer to the upper division of the university. Without minimizing the importance, however, of this so-called preparatory or transfer function for the minority of junior college students, there also has been a marked, increasing, and deserved emphasis upon a variety of terminal curriculums designed more specifically to fit the needs of the majority of students who will probably continue to complete their formal education in the junior college.

The junior college, by means of carefully organized terminal curriculums that include appropriate proportions of courses of the general or cultural type and also of the vocational or semiprofessional type, is trying to fit the young man or young woman completing them, not only to earn a better living, but also to live a better life—in the home, in the community, and as a voting citizen in the democracy of which he is to be a responsible part.

In most colleges and universities the general aspects of education are completed by the end of the sophomore year, after which intensive specialization or professional preparation receives the chief emphasis. Many students in the colleges are found not well fitted nor particularly interested in these more specialized areas; hence the large mortality by the end of the sophomore year in these institutions also, as mentioned above.

University leaders, therefore, influenced in part by their own experience and in part by the development of the junior college movement and its increasing insistence upon the importance of well-rounded general education, are urging and arranging reorganization of their institutions to give greater recognition to completion of formal education at the close of the sophomore year or the lower division for many students, and to mark this level of achievement by an appropriate recognition in the way of an academic degree.

In the past twenty years more than 100,000 young people have received the commonly recognized degree or title of associate in arts at the conclusion of a systematic course of study in the junior college. Higher educational institutions are rapidly increasing their use of the same degree, influenced in part by its extensive use in junior colleges. The associate's degree was first used in this country at the University of Chicago in 1900, but it originated, as marking the completion of a two-year college curriculum, in British universities thirty or forty years earlier.

The associate's degree, as a recognition of the successful completion of an important unit of work of college grade, has been used by the General College of the University of Minnesota since its organization in 1933. It has been employed for the completion of lower division work at Colorado College, University of Houston, Illinois Wesleyan University, University of Louisville, Birmingham–Southern College, and many other universities and liberal-arts colleges.

The most recent and most important action in this respect is that of the University of California, which has had particularly close relationships with the junior colleges of California, the state in which these institutions have had their most significant development, particularly in their terminal and semiprofessional aspects. Acting on the recommendation of

junior college and state college officials, the University of California in 1941–42 voted to confer hereafter the degree of Associate in Arts on all of the 5,000 to 10,000 students who each year complete their lower division work in its principal colleges both at Berkeley and at Los Angeles. "The purpose of the degree," the University says in its formal announcement of the new plan, "is to accentuate the function of the first two years of college in general education rather than specialized education and to give to those who withdraw from college at this point some significant recognition of the course of studies they have completed."

In February, 1942, the University of Nebraska announced the organization of a series of two-year terminal curriculums, also to be marked by the Associate in Arts or by the Associate in six other fields. In Connecticut in 1941 the four state teachers colleges were authorized to grant the degree of Associate in Science to those who complete its first two years of general work, for much the same reasons stated in the case of the far-distant University of California. A similar plan is under consideration by the teachers colleges in other states.

Thus we find in all types of educational institutions above the high school level—universities, liberal arts colleges, teachers colleges, and junior colleges in North, South, East, and West, from New England to California—a marked and increasing tendency to consider the first two college years as a distinct and significant educational unit in themselves, a suitable and desirable terminal point for the formal education of many students. We also find a rapidly accelerating tendency to recognize this degree of educational achievement by a distinctive two-year or junior college degree, the associate's degree, analogous to the use of the bachelor's degree for the completion of a four-year collegiate course of study.

The junior colleges and the senior colleges and universities increasingly are recognizing their obligation and are endeavoring to cooperate more effectively in the common educational task of giving the best possible general education to young people of both sexes before the majority of them enter the industrial, the commercial, the civic, and the home life of the nation, under normal conditions at about twenty or twenty-one years of age.

This statement was written by W. C. Eells, executive secretary of the American Association of Junior Colleges.

Higher Education for Negroes.—C. E. Rochelle completed at the University of California in August, 1942, a doctorate dissertation on "Graduate and Professional Education for Negroes." The purpose of this study was to determine: (1) the needs for graduate and professional instruction for Negroes in the states having separate schools for colored and white students; (2) the status of graduate and professional instruction for Negroes in these states; (3) the implications of the Supreme Court decision in the Gaines Case in regard to graduate and professional instruction for Negroes in the Southern states; (4) the future possibilities for graduate and professional instruction for Negroes in the Southern states.

The data for this investigation were collected by personal interviews, and first-hand examination of the following: catalogues and bulletins of Negro institutions offering graduate and professional instruction; school codes and constitutions of

the Southern states; reports and proceedings of school officials and organizations interested in providing graduate and professional instruction for Negroes in the Southern states; surveys of higher education for Negroes in Alabama, Georgia and North Carolina; government publications; court decisions of cases relating to graduate and professional education for Negroes; and articles and studies concerning graduate instruction and professional education for Negroes.

When analyzed and organized the available data tended to show: that an increase in the facilities for graduate and professional instruction for Negroes in the Southern states is justified on account of the increase in both enrollment and graduates from Negro secondary schools and colleges; the demand for more and better qualified Negro teachers at all levels in the Southern states; and the need for a greater number of Negro physicians, dentists, lawyers, and leaders in these states.

In the Southern states Negroes receive educational opportunities on the graduate and professional levels that are neither substantially equal in degree or kind to those provided for white students. It appears doubtful that the Southern states in the near future will break down the social tradition of segregation and admit Negroes upon equal terms with white students to the public-supported graduate and professional institutions now attended only by white students.

The Supreme Court decision in the Gaines Case has focused attention upon the problem of providing graduate instruction and professional education for Negroes in the Southern states and has created an unusual interest in the solution of the problem. The development of graduate and professional departments as additions to the state colleges for Negroes appears to be the most probable method that will be used by Southern states to meet the requirements of the Supreme Court decision in the Gaines Case.

Growth Studies in 4-H Club Work.—Recent evaluation studies indicate that boys and girls in 4-H Clubs are developing in the direction of the educational objectives of the work. The third, fourth, and fifth of these studies were conducted in the dairy project in Massachusetts and the sheep and clothing projects in Missouri.

Tests for some objectives of these projects were prepared and given to members, and in some studies to non-members, at the beginning of the project year and again at the end.

The members averaged less than fourteen years of age. They made gains in acquisition of useful information relating to the aspects of farming and homemaking, and in other objectives tested.

The girls in the 4-H Clothing Clubs of six Missouri counties understood patterns better at the end than at the beginning of the project. They improved in their ability to select materials, designs, colors, and accessories for dresses appropriate to the occasion and the girl. Confidence in their own ability to make garments increased considerably. They developed greater appreciation for the clothes they made for themselves. The girls were more adept in clothing skills at the end of the project than at the beginning.

The results of the tests showed that the 4-H sheep club members of eleven Missouri counties learned practical information relating to the care, feeding, and marketing of sheep. The importance of

their project goals shifted somewhat during the project. Although the vocational goal remained in first place, the companionship and leadership goals increased in importance. The profit and organization goals decreased. The competitive goal was next to last in importance of the nine goals tested.

The members of the 4-H dairy clubs of nine Massachusetts counties made much greater gains than the equivalent group of non-members in their knowledge of practical dairy information. The members' desire to continue in school remained about the same but the non-members' desire dropped. Vocational plans of the members became more definite than the non-members during the period of the study. The differences between the equivalent groups in the last two objectives, however, were not statistically significant.

This summary was prepared by Fred P. Frutchey, Division of Field Studies and Training, Extension Service, U. S. Department of Agriculture.

Rockefeller Foundation.—During 1941 the appropriations of the Rockefeller Foundation amounted to $9,313,964. The income of the Foundation from investments during the year was $8,734,992. This income was supplemented by a transfer of $600,000 from the Principal Fund.

The appropriations were distributed for the most part in six major fields, roughly as follows:

Public health	$2,450,000
Medical sciences	2,120,000
Natural sciences	1,271,000
Social sciences	1,227,000
Humanities	1,020,000
Program in China	154,000

Of the money spent during the year, 74 per cent was for work in the United States and 26 per cent for work in other countries.

General Education Board, 1941.—Appropriations by the General Education Board during the year 1941 totaled $2,690,578. Of this sum $1,658,848 was taken from the principal fund and $1,031,730 from income. The appropriations in terms of general classifications were as follows:

Southern education		
White	$ 753,785	
Negro	1,101,720	$1,855,505
General education		351,600
Surveys, studies, and conferences		30,000
Administration		
Educational divisions	$120,874	
General	107,624	228,498
		$2,465,603
Authorizations for later appropriation		224,975
		$2,690,578

As of December 31, 1941, the unappropriated asests of the General Education Board, in terms of market value, amounted to $12,884,210.37.

Total appropriations from 1902, when the Board was founded, to December 31, 1941, have been as follows:

From principal	$152,816,053.09
From income	120,333,050.42
	$273,149,103.51

The General Education Board, founded in 1902, began its work in the South. The scope of its grants soon became national and remained on that basis until the end of the year 1940. The year now under review marks a return to primary concentration in the southern states. Without closing the door to the consideration of matters of exceptional importance

which may arise elsewhere, the trustees have decided that the limited funds now remaining at the Board's disposal shall be applied chiefly in aid of education in the South.

Projects of the American Library Association.—The 1942–43 publishing season of the American Library Association promises to bring to teachers, students, and librarians a number of important additions to educational literature.

One of the first titles that will appear —*Experimenting Together: The Librarian and the Teacher of Music,* by Josephine Dillon and Esther L. Bohman—describes the cooperation of librarian, music teacher, and pupils at Mount Auburn Elementary School in Cleveland. Several librarians who have read it praise the freshness of its approach, and in North Carolina the state supervisor in charge of music and a member of the state music committee have read it and recommended its insertion in the state course of study. This title is the third in the *Experimenting Together* series, which pictures pupils, teachers, and librarians working together in the development of dynamic educational programs. The earlier tides were: *The Librarian and the Teacher of English* (1938) and *The Librarian and the Teacher of Science* (1942). Both of these describe coordinated activities at the high school level.

Two important aids for teachers and librarians are a new editor of *Subject Index to Books for Primary Grades* and a supplement to *Subject Index to Books for Intermediate Grades.* Both of these books were prepared after a thorough study of units actually taught in schools, and provide a subject approach to material up through the sixth grade. The books indexed include nearly 2,000 carefully chosen titles.

At the college level A.L.A. will publish a *Catholic Supplement* to the Shaw *List of Books for College Libraries.* The original "Shaw list" and its supplement (1940) have been buying guides for college libraries for several years. This latest supplement will give Catholic institutions a list emphasizing their own special needs.

For the research worker, the new *Subject Guide to Reference Books* provides a practical approach to the maze of purely informational reference books. By means of 240 useful subject "units", more than 600 commonly used reference books are correlated and listed where they do the most good. A new book on *Reference Work* by Hutchins is planned to offer guidance in rendering reference service in school, college, and public libraries.

Two important meetings of the A.L.A.'s annual conference in 1942 were devoted to a study of U. S. and foreign government publications bearing on World War II. The papers of these meetings have been prepared for publication by the A.L.A. Public Documents Committee, as *Public Documents and World War II.* The appearance of this book will make available to libraries and research workers much needed data on the publications of war-time government agencies.

A much-needed handbook, *Glossary of Library Terms,* is also scheduled to appear this winter. In it research workers will find information useful in dealing with reference books, periodicals, and bibliographies.

This information was made available by the executive offices of the A.L.A. at Chicago.

Suggested Studies in Adult Education.—The Institute of Adult Education of Teachers College, Columbia University, was established October 1, 1941, for the purpose of conducting an intensive study of the opportunities, problems, materials, and methods of adult education, and for assistance in the training of leaders and workers in the field. From the start, it was contemplated that many of the resources of the Institute would be devoted to the preparation and publication, on an experimental basis, of materials of instruction for adults at various educational levels. Such plans included the issuance of pamphlets, brochures, discussion guides, and materials for educational use in connection with motion pictures and the radio. A program of basic research along sociological, psychological, and educational lines was also contemplated. This research was to be closely coordinated with the work of the Institute and carried on both by the appropriate divisions at Teachers College and by the Institute itself.

The task of preparing a tentative program for the newly established Institute was commenced on November 17, 1941, immediately following the appointment by Dean William F. Russell of nine Associates of the Institute, chosen from the faculty of Teachers College. The Associates, it was agreed, would serve as an advisory and policy-determining body, the administrative function being the responsibility of the Executive Officer of the Institute, Professor Morse A. Cartwright, assisted by Miss Dorothy Rowden, Associate in Adult Education. The Institute is supported by a separate grant made by the Carnegie Corporation of New York, to be expended over a period of ten years from October 1, 1941, but administratively it is an integral part of Teachers College and, as such, its Executive Officer is answerable to the Trustees of Teachers College, reporting through the Dean of the College.

The first bulletin of the Institute, a twenty-six page booklet entitled *Suggested Studies in Adult Education,* is the result of the deliberations of the Associates of the Institute.

Committee on Secondary Education.—The Consultative Committee on Secondary Education was appointed in December, 1941, by the National Committee on Coordination in Secondary Education. Its function is that of being advisory on general problems in secondary education to the U. S. Office of Education and to other national agencies that may desire its services. Very appropriately at this time its first report is on *Secondary Schools and the War Effort.* The secretary is C. A. Jessen of the Office of Education.

Education Abstracts.—With the May–June, 1942, issue, the editorship of *Education Abstracts,* for nearly seven years the only publication containing critical summaries of current educational books and articles, has been assumed by William W. Brickman and associates. He replaced Paul M. Cook, executive secretary of Phi Delta Kappa, who served from 1937. The journal was started in 1936 by Norman J. Powell, the first editor. The editorial and business offices are located at 32 Washington Place, New York, N.Y.

Education Abstracts publishes abstracts, reviews, and brief notices of books, magazine articles, pamphlets, and other publications in the field of education and in closely related areas. The abstracts are

50–300 words long, furnish complete bibliographical data, including the identity of the author in many cases, and frequently contain critical and evaluative comment.

There are about forty subdivisions covering the entire educational area. Each subdivision, e.g., administration, educational psychology, guidance, religious education, etc., is followed by cross references. Every issue contains an author index. In addition, there is an annual author index.

The staff of abstractors is drawn from all parts of the country and, before the War, from abroad. Practically every level of educational activity and interest is represented in the group. The members of the co-operating editorial board occupy leading positions in the educational world.

The editors aim, as have their predecessors, to make *Education Abstracts* a highly useful tool to the research worker, whether graduate student, professor or public school officer, by enabling him to obtain descriptions and evaluations of current educational literature. It can be used as a supplement to such research and bibliographic aids as *Education Index*, *Review of Educational Research*, and Monroe's *Encyclopedia of Educational Research*. Librarians, administrators, and teachers will find *Education Abstracts* helpful in keeping up-to-date.